WALTER AND ANDRÉE DE NOTTBECK FOUNDATION
SCIENTIFIC REPORTS

No. 40

Linking taxonomy and environmental 18S-rRNA-gene sequencing of Baltic Sea protists

MARKUS MAJANEVA

Academic dissertation in Aquatic Sciences,
to be presented, with permission of
the Faculty of Biological and Environmental Sciences
of the University of Helsinki,
for public criticism in Economicum Luentosali,
Arkadiankatu 7, Helsinki,
on May 24th 2013, at 12 noon.

HELSINKI 2013

This thesis is based on the following papers, which are referred to by their Roman numerals:

I. **Majaneva, M.**, Remonen, I., Rintala, J.-M., Belevich, I., Kremp, A., Setälä, O., Jokitalo, E. & Blomster, J.: *Rhinomonas nottbeckii* n. sp. (Cryptomonadales) and the evidence for paraphyly of the genus *Rhodomonas*. – Manuscript.

II. **Majaneva, M.**, Rintala, J.-M., Hajdu, S., Hällfors, S., Hällfors, G., Skjevik, A.-T., Gromisz, S., Kownacka, J., Busch, S. & Blomster, J. 2012: The extensive bloom of alternate-stage *Prymnesium polylepis* (Haptophyta) in the Baltic Sea during autumn–spring 2007–2008. –European Journal of Phycology 47: 310-320.

III. Rintala, J.-M., Hällfors, H., Hällfors, S., Hällfors, G., **Majaneva, M**. & Blomster, J. 2010: *Heterocapsa arctica* subsp. *frigida* subsp. nov. (Peridiniales, Dinophyceae) – description of a new dinoflagellate and its occurrence in the Baltic Sea. – Journal of Phycology 46: 751-762.

IV. **Majaneva, M.**, Rintala, J.-M., Piisilä, M., Fewer, P.D. & Blomster, J. 2012: Comparison of wintertime nano-sized eukaryotic communities in the Baltic Sea ice and water, based on sequencing of the 18S rRNA gene. – Polar Biology 35: 875-889.

V. **Majaneva, M.**, Hyytiäinen, K., Rintala, J.-M., Müller, S., Majaneva, S., Nagai, S., Motoshige, Y., Fujiwara, A., Varvio, S.L., Autio, R. & Blomster, J.: Sympagic organisms in drift and fast ice samples of Gulf of Finland (Baltic Sea): based on microscopic observations and 454 sequencing of ribosomal 18S gene. – Manuscript.

The research papers have been reproduced with the kind permission of Taylor & Francis (II), John Wiley and Sons (III) and Springer Science and Business Media (IV).

CONTRIBUTIONS	I	II	III	IV	V
Original idea	MM	JR, SeH, SuH, MM	GH, SeH, JR	JB	JR, MM
Study design	MM, JR, JB	MM, JR, SuH	JR, GH, SeH, HH, MM	JB, JR, DF, MM	JR, MM, RA
Culturing	MM	-	JR	-	-
Light microscopy	MM	SuH, SeH, AS, SG, JK, SB	JR, GH, SeH	KK	JO
Electron microscopy	JR, IB, MM	JR	JR	-	-
DNA extraction, PCR	MM, IR	MM	JB	MP, MM	MM, AF, YM
Molecular data analysis	MM, IR	MM	JB, MM	MM	MM, KH
Responsible for ms	MM	MM, SuH	JR, HH	MM, JB	MM
Contributions to ms	IR, JR, IB, OS, AK, JB	JR, SeH, AS, SG, JK, JB	MM, GH, SeH, JB	JR, MP, DF, JB	JR, SMü, SMa, JB

AF = Atushi Fujiwara, AK = Anke Kremp, AS = Ann-Turi Skjevik, DF = David P Fewer, GH = Guy Hällfors, IB = Ilya Belevich, IR = Iina Remonen, JB = Jaanika Blomster, JK = Janina Kownacka, JO = Johanna Oja, JR = Janne-Markus Rintala, KH = Kirsi Hyytiäinen, KK = Kai Kivi, MM = Markus Majaneva, MP = Maria Piisilä, OS = Outi Setälä, SB = Susanne Busch, SeH = Seija Hällfors, SG = Slawomira Gromisz, SMa = Sanna Majaneva, SMü = Susann Müller, SuH = Susanna Hajdu, YM = Yasuike Motoshige

Supervised by Adjunct Professor Jaanika Blomster, University of Helsinki, Finland; Adjunct Professor Harri Kuosa, Finnish Environment Institute, Finland

Reviewed by Dr Rebecca J. Gast, Woods Hole Oceanographic Institute; Professor David A. Caron, University of Southern California

Examined by Associate Professor Tove M. Gabrielsen, The University Centre in Svalbard

LINKING TAXONOMY AND ENVIRONMENTAL 18S-rRNA-GENE SEQUENCING OF BALTIC SEA PROTISTS

MARKUS MAJANEVA

Majaneva, M. 2013: Linking taxonomy and environmental 18S-rRNA-gene sequencing of Baltic Sea protists. – W. & A. de Nottbeck Foundation Sci. Rep. 40: 1–49. ISBN 978-952-67851-3-4 (paperback), ISBN 978-952-67851-4-1 (PDF).

Protists are unicellular eukaryotes. Some protistan species may be impossible to distinguish under the light or even electron microscope, and a complete balanced study of protistan taxonomy requires molecular analysis and light and electron microscopy. One of the main applications of taxonomic work is the assessment of diversity of organisms in an ecosystem. However, uncertainty in taxonomic precision undermines the diversity measures. DNA sequence data provide assistance since they are easily transformed to numbers that can be compared systematically and in a similar way throughout the eukaryotic domain, using sequence similarity to define operational taxonomic units (OTUs). DNA-based assessment of diversity is called environmental sequencing. The most commonly used gene in the environmental gene sequencing for eukaryotes and also in the protistan taxonomic studies has been the small subunit (18S) ribosomal RNA gene of the ribosomal operon. Also, internal transcribed spacers (ITS) are used.

The studies in this thesis were conducted with Baltic Sea protists. The Baltic Sea is a sub-arctic brackish-water basin that partially freezes over every winter. If the salinity of parent water is higher than 0.6, the forming ice has a semi-solid structure with solid ice crystals and saline water (brine) channels. The brine channels offer habitats for small-sized organisms. Due to the low salinity of the Baltic Sea, the brine channels are small, and therefore, the Baltic Sea ice eukaryotic community is dominated by protists. Studies on Baltic sea-ice biology have been accumulating since the 1980's, but there are still gaps on knowledge; for example, what protistan species and how many there are associated with sea ice. In this thesis, morphological, molecular and ecological information was combined to delineate species of a Baltic Sea cryptomonad, haptophyte and dinoflagellate. Protistan community composition in Baltic Sea ice was assessed with environmental sequencing, and diversity estimates were compared in different types of ice. The taxonomic and environmental-sequencing studies were linked using the gathered taxonomic information to evaluate the accuracy of the molecular diversity-measurement method.

A new cryptomonad species, *Rhinomonas nottbecki*, was described based on morphological characters distinguished by light- and electron-microscopy together with molecular evidence from 18S rRNA gene and ITS region. The same approach was applied to the identification of the alternate stage *Prymnesium polylepis* (Haptophyta), which bloomed in the whole Baltic Proper during autumn–spring 2007–2008. Also, a novel cold-water and sea-ice associated dinoflagellate subspecies, *Heterocapsa arctica* subsp. *frigida*, was described. Environmental 18S-rRNA-gene sequencing revealed that the richest eukaryotic lineages inhabiting the Baltic Sea ice were ciliates, cercozoa, dinoflagellates and diatoms. The

different developmental stages and types of ice had different community composition. Protistan richness was higher in ice than water even though water included more divergent lineages. The Baltic fast ice had higher richness than pack and drift ice.

The results of this thesis showed that there remains novelty to be described in the Baltic Sea, and what we know about the protistan community in Baltic Sea ice now is very incomplete. Although the environmental sequencing produced data that met the requirements of calculation of comparable diversity indices (all taxa defined at the same level), revealed cryptic taxa, and gave higher protistan richness than basic light microscopy of fixed samples, the lack of taxonomic detail was not restricted to the light-microscopic surveys but was also a result of the environmental-sequencing approach. This was shown when the environmental-sequencing approach was applied on the 18S-rRNA-gene data of the cryptomonad family Pyrenomonadaceae and the haptophyte genus *Prymnesium*. Only one Pyrenomonadaceae and two *Prymnesium* OTUs were found although both data sets included 15 distinct taxa. Errors in environmental sequencing and alignment make the use of high similarity levels in the OTU definition questionable, and the variability in the 18S rDNA is not equal within different eukaryotic lineages. Consequently, use of lower similarity level (97 %) is justifiable in the environmental-sequencing, but the approach used gave conservative estimates of the protistan richness in the Baltic Sea ice.

The overall conclusion is that we need to apply all available techniques when assessing the diversity of protists, as each technique provides a biased perspective on nature. A labor intensive taxonomic approach that includes the study of live cells by light microscopy, detailed morphological description based on electron microscopy and phylogenetic analysis of suitable genetic markers gives us the best chance of finding out how many different species of protists live within the Baltic Sea ice or any other environment, and what they do.

Markus Majaneva, Tvärminne Zoological Station, J. A. Palménin tie 260, FI-10900 Hanko, Finland, Department of Environmental Sciences, P.O. Box 65, FI-00014 University of Helsinki, Finland and Finnish Environment Institute, Marine Research Centre, P.O. Box 140, FI-00251 Helsinki, Finland.

CONTENTS

1. INTRODUCTION

1.1. Classification and principles of taxonomic studies of eukaryotic domain of life

1.1.1. Classification of Eukaryota

There are three major lineages (domains) of living organisms: Bacteria, Archaea and Eukaryota. My thesis will focus on the eukaryotes, which are cellular organisms that carry their genetic material in a membrane-bound nucleus.

The eukaryotic classification has developed dramatically since Linnæus (1753, 1758) invented the present system where species have binominal names and are hierarchically combined into higher-rank groups based on shared characteristics (synapomorphies). A synapomorphy is a character that is shared by taxa and their most recent common ancestor, whose ancestor in turn does not have that character (Page & Holmes 1998). A taxon is "A taxonomic unit, whether named or not: i.e. a population, or group of populations of organisms which are usually inferred to be phylogenetically related and which have characters in common which differentiate (*q.v.*) the unit (e.g. a geographic population, a genus, a family, an order) from other such units. A taxon encompasses all included taxa of lower rank (*q.v.*) and individual organisms." according to the International Commission of Zoological Nomenclature (1999).

First, living organisms were divided into two kingdoms: animals and plants, but a broader classification system started to emerge during the 19th century when microscopical anatomy and historical dimensions were taken into consideration, first by Ehrenberg (1830) and Dujardin (1841) and later by Haeckel (1866), Bütschli (1880–1889), and several others (reviewed in Hausmann et al. 2003) who were inspired by Darwin's (1859) theory of the origin of species. In the late 20th century, growing knowledge from DNA sequence data (e. g. Woese & Fox 1977, Sogin et al. 1986, Van de Peer et al. 1996, Kuvardina et al. 2002, Burki et al. 2012) further revolutionized the classification system, and the first two kingdoms, animals and plants, have become trivial lineages within the great bush of eukaryotic life.

Here (Table 1), I present the outline of the most recent classification system (Adl et al. 2012). Our understanding of eukaryotes' inter-relationships is still maturing and the classification of Adl et al. (2012) is not the final truth but should be viewed as a hypothesis that may be tested and revised as new knowledge accumulates.

Table 1. The classification of eukaryotes. Modified from Adl et al. (2012). Grey background indicates doubtful placement in the classification.

Domain		Supergroups	Examples	Examples
Bacteria				
Archaea				
Eukaryota	Amorphea	Amoebozoa	Mycetozoa	
			Discosea	Flabellinia
		Opisthokonta	Fungi	Dikarya
				Chytridiomycota
			Choanomonada	
			Metazoa	
			Apusomonada	
			Breviata	
		Excavata	Metamonada	
			Discoba	Euglenozoa
	Diaphoretickes	Sar	Cercozoa	
			Alveolata	Ciliophora
				Dinoflagellata
			Stramenopiles	Diatomea
				Chrysophyceae
		Archaeplastida	Glaucophyta	
			Rhodophyceae	
			Chloroplastida	Chlorophyta
				Charophyta
			Cryptophyceae	Cryptomonadales
				Kathablepharidae
			Centrohelida	
			Telonemia	
			Haptophyta	
	Incertae sedis Eukaryota			145 groups or genera

The three domains of life represent monophyletic taxonomic units. A monophyletic unit includes all descendants of an ancestral taxon (Page & Holmes 1998). According to Adl et al. (2012) the eukaryotic domain includes three larger monophyletic groups plus several groups whose position within Eukaryota is uncertain, called *incertae sedis* Eukaryota (Table 1). The first of the major groups is called Amorphea and it includes the supergroups Amoebozoa (including lineages such as Discosea and Mycetozoa) and Opisthokonta (e.g. Fungi, Metazoa and Ichthyosporea). The second is called Diaphoretickes, including the supergroups Archaeplastida (e.g. Glaucophyta, Rhodophyceae and Chloroplastida) and Sar (e.g. Stramenopiles, Alveolata and Rhizaria), but probably also some *incertae sedis* Eukaryota such as Cryptophyceae, Telonemia and Haptophyta. Diaphoretickes includes most of eukaryote lineages. The third major group has no name but includes the supergroup Excavata (e.g. Metamonada, *Malawimonas* and Discoba).

By convention, unicellular (mainly microscopic, e.g. yeast) organisms have been separated from multicellular (e.g. tiger, pine) organisms. The unicellular eukaryotes have been gathered under varying names such as Animalcula, monads, Infusoria, Protozoa and Protista (reviewed in Hausmann et al. 2003). However, this division based on the level of cellular organization is problematic since the present classification system (Adl et al. 2012) shows that it is not consistent with our understanding of evolutionary relationships. Nevertheless, Adl et al. (2005) argue that the popular term protist should be retained in use to describe unicellular eukaryotes without cell differentiation into tissues. The term protist is used in this thesis although it is not a systematic term. Instead, it is a practical term.

1.1.2. Taxonomic studies of protists

Some scientists are interested in defining entities into groups based on characters those entities share, and naming those groups. The fields in science that those people practice are called taxonomy and systematics. By one definition, taxonomy describes groups of organisms and assigns scientific names to these groups, while systematics is the theory and practice of grouping of entities into species, arrangement of species into larger groups, and naming of those groups (Judd et al. 2008). By further grouping and naming, systematics produces a classification. Thus, the Adl et al. (2012) classification is a product of taxonomic and systematic work.

Elements within classifications are based on shared (synapomorphic) morphological characters. When the relations of taxa are based on synapomorphies, the result is a hypothesis of the phylogeny of those taxa. Initially, investigators used morphological information obtained by light microscopy observations of large numbers of cells, often growing in culture. More characters were found when electron microscopes were invented in the 1930's (Rudenberg & Rudenberg 2010) and applied to protist taxonomy in the 1960's and beyond (Patterson 1999). The need to work with cells in cultures mostly limited studies to those species that are cultivable, although electron microscopy may be applied to environmental samples as well (e.g. Patterson 1985).

The implementation of sequencing of environmental samples has uncovered an array of novel taxa not found with microscopes (e.g. Not et al. 2007), and clarified phylogenetic relations of some enigmatic protists (e. g. Burki et al. 2009). Consequently, DNA sequence data has become an essential part of protistan taxonomy and phylogenetic analyses. A complete balanced study of protistan taxonomy requires molecular analysis and light and electron microscopy.

A fundamental task of taxonomy is to delineate species. Although the first impression may be that this is simple, there are severe obstacles to cross. The first, and perhaps the most difficult one, is how to define the term species. There is a vast literature (see e.g. Mayr 1957, Slobodchikoff 1976, Andersson 1990, Hausdorf 2011, Boenigk et al. 2012) dealing with this so-called species problem where tens of more or less overlapping species concepts are discussed. However, there is no species category in nature (Ereshefsky 2011, Boenigk et al. 2012). The present organisms that we see are the result of a continuing dynamic process of evolution, and 'species' are subsets of the continuum that taxonomists recognize.

Classification of protists into higher rank groups is based on synapomorphies these groups share, for example the shape of

cristae in the mitochondria, flagellar 9 + 2 axoneme structure and many others, see e.g. Patterson (1999). Within these higher-rank groups, different characters may be used for species delineation. For example, many dinoflagellates have organic body scales on their surface and an armor called theca that is composed of cellulose plates in various shapes and arrangements, and this thecal plate arrangement together with the organic body scales are used to delineate species (e.g. Iwataki et al. 2003). Likewise, the layer of unmineralized body scales is used to delineate species of haptophytes (e.g. Jensen 1998). Cryptomonads are harder to identify to species level. They can be delineated based on for example cell size and shape, but sometimes this process requires DNA sequence data (e.g. Hoef-Emden 2007). Here, I have used a practical, integrative taxonomic approach where morphological, molecular and, in some extent, ecological information was combined to delineate species.

1.2. Measurement of protistan richness and diversity using environmental gene sequencing

One of the key areas of ecology, a scientific discipline studying the relationships of living organisms and their environment, is to assess the diversity of organisms in an ecosystem. The term diversity has several meanings. In biology, it is "Most simply, the species richness of a community or area, though it provides a more useful measure of community characteristics when it is combined with an assessment of the relative abundance of species present" (Allaby 2010). The United Nations Environment Programme (1992) defines (biological) diversity as followed: '"Biological diversity" means the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic systems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems.', which Harper & Hawksworth (1995) corrected by changing ecosystem (includes physical environment) to community (includes only living organisms). The term biodiversity is a contracted form of biological diversity (Wilson 1988). According to Harper & Hawksworth (1995), biological diversity may refer to genetic (within species), organismal (between species or higher rank taxa; number of taxa) and ecological (community level; number of trophy levels) diversity. Here, I use the word diversity according to Harper & Hawksworth (1995), meaning organismal diversity, and hence, diversity refers to taxon richness (number of taxa) and abundance of those taxa (number of individuals) in one or several samples.

Another issue is how to measure this diversity (see Magurran 2004). It may seem simple to count the species number of a sample, but identifying protists to species level is laborious and requires special expertise that has been gathered through years of microscope work. Even so, as mentioned above, protistan species may be impossible to distinguish under the light or even electron microscope (Lowe et al. 2011). The resulting uncertainty in taxonomic precision undermines the diversity measures (Wu 1982). One way of dealing with the uncertainty is to use surrogates of species (Magurran 2004). For example, Lee (1997) shows that genus- and family-level richness estimates the underlying species richness pretty well, but each taxon and system must be verified case by case (Magurran 2004).

DNA sequence data provide assistance since they are easily transformed to numbers that can be compared systematically and

in a similar way throughout the eukaryotic domain. The level of comparison can be chosen by the investigator depending on the purpose of the analysis. For example, sequences are grouped based on their similarity: an investigator may discover that most different species have approximately 1 % difference in their gene sequence used for comparison, and so the investigator groups sequences into groups with 99 % similarity. However, the interspecies variation in genes may differ broadly (e.g. Caron et al. 2009) depending on the evolutionary rates within each lineage of eukaryotes, and there is no universally correct level of similarity. Therefore, groups formed on the basis of similarity of sequences are called operational taxonomic units (OTUs), not species (Sokal & Sneath 1963, Blaxter et al. 2005).

To measure diversity of protists using this DNA-based approach, an investigator will collect samples from nature, extract DNA (or RNA) from the samples and work out the base sequence of the selected gene present in the samples. This approach is called metagenomics or environmental gene sequencing (Chen & Pachter 2005). The gene sequences are then grouped together according to their similarity at a constant level, and diversity measures can be calculated in a consistent way across the eukaryotic domain. The result does not rely on the accuracy of species determination skills of the investigator.

The investigator can compare the number of taxa in different samples (richness) or calculate different diversity indices. These indices can be based on incidence or abundance of taxa (Magurran 2004). The incidence-based indices take into account only the information of presence and absence of taxa in several samples (e.g. Lee & Chao 1984, Chao 1987); while the abundance-based indices take into account both the presence-absence and abundance of taxa in samples. The latter indices incorporate a measure of evenness or how equally abundant each taxon is in the samples (e.g. Shannon 1948, Simpson 1949, Chao 1987). However, the use of abundance-based indices in environmental sequencing is problematic since the abundance of gene sequences is not linearly related to the abundance of taxa but is proportional to the gene-copy number in each taxon (Prokopowich et al. 2003).

A slightly different approach measures the diversity based on the divergence of the gene sequences. Lozupone & Knight (2008) suggested that these divergence-based indices may be the most suitable when comparing diversity of environmental-sequencing samples. They are based on a phylogenetic tree constructed from the DNA sequences and not on the number and/or abundance of taxa. A phylogenetic tree is a mathematical structure modeling the evolutionary history of a group of sequences or organisms (Page & Holmes 1998). The indices may represent the phylogenetic richness of the samples, using a sum of the lengths of branches (e.g. Faith 1992) or the phylogenetic evenness, using the average divergence between two randomly chosen sequences in the samples (e.g. Martin 2002). This approach circumvents the artificial grouping of sequences based on their similarity; instead, it measures the evolutionary richness or diversity of the community, and may result in quite different view on the community than the taxon-based indices (Lozupone & Knight 2008).

1.3. The ribosomal gene operon and gene trees

The most commonly used gene in the environmental gene sequencing of eukaryotes (e.g. Moon-van der Staay et al. 2001) and

also in the protistan phylogenetic analyses (e.g. Van de Peer et al. 1996) has been the small subunit (18S) ribosomal RNA (rRNA) gene. It is a part of the ribosomal gene operon (a cluster of genes regulated and transcribed together) that codes for ribosomal RNAs, which in turn are part of ribosomes. Ribosomes are assemblies of rRNAs and proteins and translate messenger RNAs into proteins. The ribosomal gene operon includes a nontranscribed spacer (NTS), an external transcribed spacer (ETS), the 18S gene, an internal transcribed spacer 1 (ITS1), a 5.8S gene, an ITS2 and a 28S gene.

The operon is usually repeated multiple times in eukaryotic genomes (Prokopowich et al. 2003). Ribosomal genes are highly expressed because of the central role of ribosomes in the functioning of every organism; ribosomes assemble proteins. Thus, a low amount of material is needed for both DNA and RNA sequencing. In addition, the 18S rRNA gene includes highly conserved and variable nucleotide-sequence regions allowing phylogenetic reconstruction and organism recognition at various taxonomic levels (Pawlowski et al. 2012). The 18S rRNA gene has been shown to be a useful marker to distinguish species in some (e.g. Zimmermann et al. 2011), but not in all (Pawlowski et al. 2012) groups. As an alternative, the more variable ITS regions have been used (e.g. Hoef-Emden 2007). Also, ribulose-bisphosphate carboxylase (*rbc*L) genes (e.g. Takishita et al. 2000) and other protein coding genes (e.g. Edgcomb et al. 2001, Parfrey et al. 2010) have been used independently or included in analyses allowing more comprehensive phylogenetic analyses.

The relationship between the ribosomal gene sequences of the investigated taxa can be visualized using phylogenetic gene trees. The tree is constructed using the base sequence of the gene as characters. The resulting tree is a hypothesis of the evolutionary history of the characters used, not species.

The sequences are aligned so that homologous (two similar characters that are both inherited from their ancestor which also had that character, Page & Holmes, 1998) bases are compared. However, several changes for example from base A to T to G and back to A at one particular base position may have occurred. These changes are estimated by different evolutionary models, for example the general time reversible model (GTR, Tavaré 1986).

The phylogenetic trees can be built based on distances between sequences or based on the bases directly. Another division of the tree-building methods is based on how they construct trees. Clustering methods use an algorithm to add new sequences to an initial tree. Optimality methods use optimality criteria to choose the best tree among the set of all possible trees. Optimal trees are ranked based on the relationship between tree and data using an evolutionary model like GTR (Page & Holmes 1998). An example of a tree-building method that uses distances and clustering algorithm is neighbor joining. Maximum parsimony and maximum likelihood methods are examples of methods that operate directly on the sequences and use an optimality criterion to choose the best tree. The best tree is the one that requires fewest evolutionary changes (maximum parsimony), or the one that is most likely to have produced the data (maximum likelihood) (Page & Holmes 1998). Maximum likelihood trees can be searched within the Bayesian framework where the likelihood function is combined with prior probability (Huelsenbeck et al. 2001).

To judge the strength of support of the constructed tree, Felsenstein (1985) suggest

the use of bootstrapping. Bootstrapping generates a pseudoreplicate from the sequence data by sampling with replacement, and a tree is built based on the pseudoreplicate. This process is repeated multiple times, usually 100–1 000 times. The bootstrap value is the percentage of occurrence of a certain clade in all the trees. Within the Bayesian framework, the strength of support is based on posterior probabilities that represent the probability that the corresponding clade is true given the used evolutionary model, the prior probability, and the data (Huelsenbeck et al. 2002).

Information from different tree building methods and support values are drawn into one representative tree called consensus tree (Page & Holmes 1998).

1.4. Baltic Sea and sea ice

The Baltic Sea is a sub-arctic (midpoint 60°N, 20°E) brackish-water basin that has undergone several freshwater and brackish-water phases caused by land uplift and rise in global sea-level (Tikkanen & Oksanen 2002, Leppäranta & Myrberg 2008) after the end of the Weichselian glaciation 15 000–15 500 years ago (Björck 1995). The mean salinity of the Baltic Sea is 7.4 (Leppäranta & Myrberg 2008) with a salinity gradient from Bothnian Bay (salinity <1) to Skagerrak (20–24). The salinity gradient is due to large input of freshwater from numerous rivers and the shallow and narrow connections to the Atlantic Ocean via the Danish Straits. The mean depth of the Baltic Sea is 54 m and it covers an area of 393 000 km^2 (Leppäranta & Myrberg 2008). Stemming from the short history of the Baltic Sea and its brackish water, the macroscopic species richness of the Baltic Sea is low with only few endemic species (Remane 1934, Pereyra et al. 2009), and there is still need for information on protist diversity.

Because of its northern situation, the Baltic Sea partially freezes over every winter. The ice covers on average about 40 % of the sea (Granskog et al. 2006). The ice-covered period in the northern Bothnian Bay may be longer than the ice-free season, while in the central Baltic Sea, ice is formed only during severe winters and does not persist for more than a few weeks.

If the salinity of parent water is higher than 0.6, the forming ice has a semi-solid structure (Palosuo 1961, Petrich & Eicken 2010) with solid ice crystals and channels and pockets of liquid water with concentrated dissolved constituents. During the freezing process, the dissolved constituents of the parent water are removed from the crystallizing ice to the remaining water called brine, which salinity may rise up to 173 in the Antarctic (Kottmeier & Sullivan 1988) but hardly over 50 in the Baltic Sea (Meiners et al. 2002).

The brine within the channels and pockets of ice offer habitats for small-sized organisms. The diameter of the channels is from micrometer to several centimeters (Eicken et al. 1995). The geometry and volume of the brine channels correlates with salinity of the parent water and temperature. The habitable area within the ice is substantially smaller in low-saline seas (e.g. the Baltic Sea) and in low temperature during winter than in truly marine seas and near zero temperature during spring. Due to the small size of the brine channels, the Baltic eukaryotic community is dominated by protists.

Knowledge on Baltic sea-ice biology has been accumulating since the 1980's with publications by e.g. Huttunen & Niemi (1986), Vørs (1992), Norrman & Andersson (1994), Ikävalko & Thomsen (1996, 1997), Ikävalko (1998), Haecky & Andersson (1999), Meiners et al. (2002), Kaartokallio

et al. (2007), Rintala et al. (2006, 2010) and Piiparinen et al. (2010). The sea ice may have an effect on the community composition of the major annual pelagic biomass peak in spring (Kuosa et al. 1992, Haecky et al. 1998) and the sea-ice community may be a source of energy for water column food webs (Tamelander et al. 2008) and therefore, contribute reasonably to the primary and secondary production of ice-covered areas (Mikkelsen et al. 2008). However, there are still gaps in our knowledge on how many and which protistan species there are associated with the Baltic Sea ice.

2. STUDY QUESTIONS

My thesis is divided into two parts: taxonomical studies of some Baltic Sea flagellated protists (papers **I-III**) and environmental 18S-rRNA-gene sequencing of Baltic Sea ice and water samples (papers **IV-V**). The aim of my thesis is to link these two parts to gain new knowledge on the protists inhabiting the Baltic Sea water and ice. I have used both morphological and molecular evidence when studying taxonomy of a cryptomonad (**I**), a haptophyte (**II**) and a dinoflagellate (**III**). Diversity of protists in Baltic Sea ice was investigated using molecular methods combined with some morphological information (**IV-V**). My study questions were:

1) Is the newly collected cryptomonad a novel species? (**I**);
2) What haptophyte species was blooming during autumn–spring 2007–2008? (**II**);
3) Is the cultured dinoflagellate a species or a sub-species? (**III, this thesis**);
4) What is the composition of the protistan community in Baltic Sea ice? (**IV-V**);
5) Do molecular protistan community diversity estimates differ in different types of ice? (**IV-V**);
6) Evaluation of the accuracy of the molecular diversity-measurement method using the gathered taxonomic information (**this thesis**).

3. MATERIAL AND METHODS

3.1. Cultures (I, III)

Cultured strains of flagellated protists (**I, III**) were isolated from the Hanko Peninsula area or obtained from culture collections (CCAP, CCMP, CPCC, NIES and SCCAP). The strains isolated from the Hanko Peninsula area were described as new taxa *Rhinomonas nottbecki* sp. inedit. (**I**) (hereafter referred as *Rhinomonas nottbecki*, the taxon will be published independently, and the occurrence of the name in this thesis is for convenience only and has no standing in nomenclature) and *Heterocapsa arctica* subsp. *frigida* (**III**). The novelty of those taxa was made sure by comparing them to the earlier species descriptions and to the strains obtained from culture collections.

Rhinomonas nottbecki cells were isolated from water samples collected from Storfjärden in June 2007 and 2008 by Anke Kremp and Outi Setälä (**I**, Fig. 1). Single cryptomonad cells were isolated by micropipette from size-fractionated (10-20 µm and <10 µm) or concentrated (reverse filtration) samples into culture wells of 24-well tissue culture plates containing f/8 (-Si) medium (Guillard 1975) of aged and autoclaved Baltic Sea seawater (salinity 6). Actively growing clonal cultures were transferred to vented 50 ml polycarbonate tissue culture flasks (Greiner Bio-One, Kremsmünster, Austria) and maintained in f/2 (-Si) culture medium at

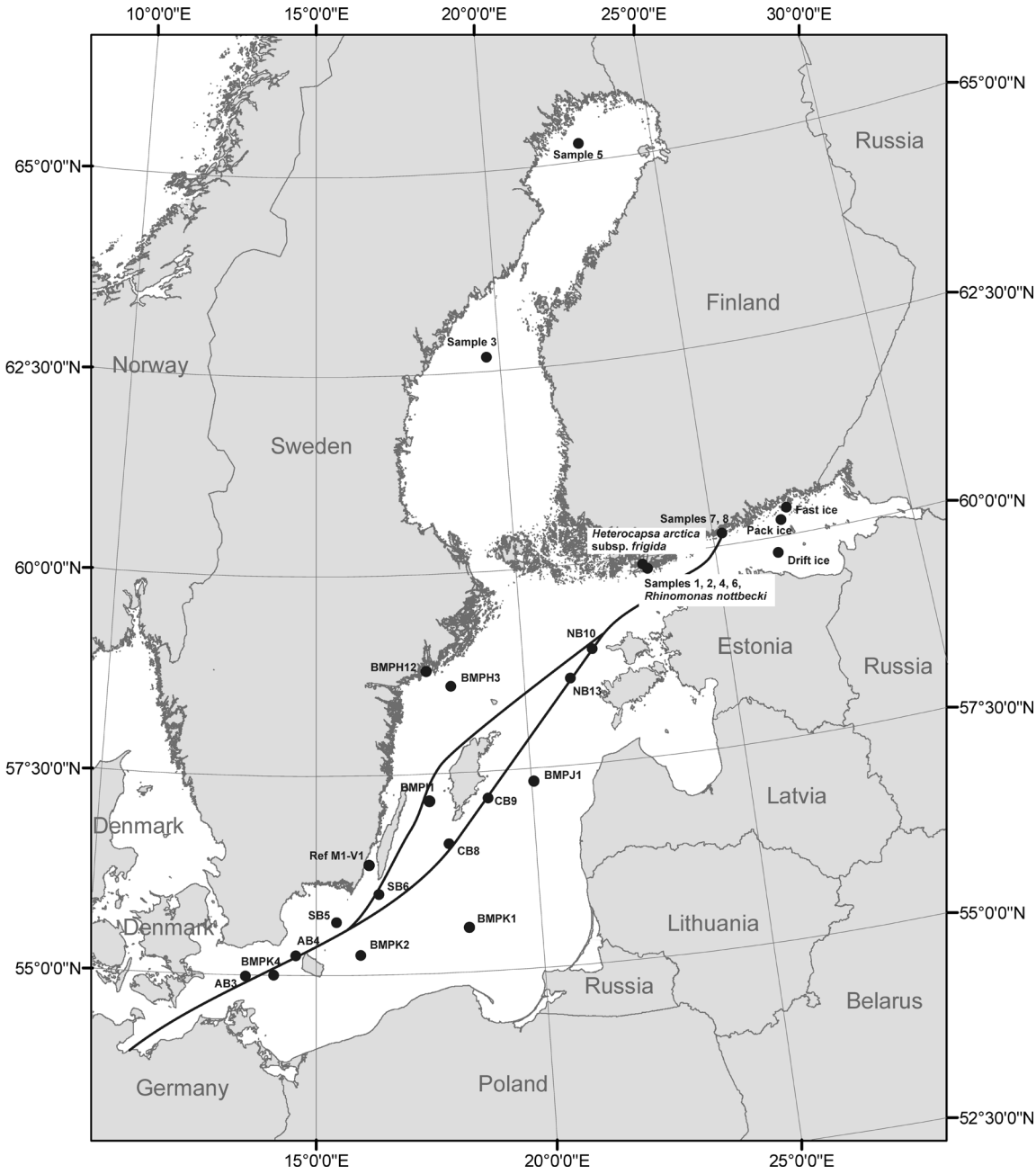


Fig. 1. A map showing the sampling sites within this study. The sampling locations where cultured strains (**I, III**) were collected are named accordingly. *Prymnesium polylepis* samples (**II**) are named in style AB3. The line denotes the approximate route of the Alg@line ships where *Heterocapsa arctica* subsp. *frigida* samples (**III**) were taken. The community samples are marked either with sample number (**IV**) or with the ice type (**V**).

4 °C and 16 °C with 12:12 light:dark cycle, and 100 μE m^{-2} s^{-1}.

Motile *Heterocapsa arctica* subsp. *frigida* cells were isolated from sea ice collected in Santala Bay in March 2001 by Janne-Markus Rintala (**III**, Fig. 1). Each cell was isolated using autoclaved glass Pasteur pipettes and put into a 50 ml Cellstar® tissue culture flask (Greiner Bio-One) filled with f⁄2 (-Si) medium. The strain was kept at 4 °C under 40 μE m^{-2} s^{-1} light with a daily light:dark cycle of 8:16.

The strains obtained from culture collections were grown in above described medium that had artificially (Instant Ocean® Sea salt, Spectrum Brands, Madison, WI, USA) raised salinity of 30 (**I**) or 35 (**III**) at 4 or 16 °C. The strain CPCC344 was grown in f/2 (-Si) medium of fresh, filtered (0.2 μm) tap water at 16 °C.

3.2. Sampling (II-V)

3.2.1. Monitoring samples (II-III)

The countries around the Baltic Sea carry out monitoring that is coordinated by the Baltic Marine Environment Protection Commission (HELCOM). The sampling is based on either integrated or ship-of-opportunity samples. The integrated samples (**II**) (Fig. 1) were obtained with a hose from 0–10 m or 0–20 m at the given stations according to HELCOM monitoring guidelines (HELCOM 2008) from October 2007 to May 2008 by the HELCOM partners. The ship-of-opportunity samples (**II-III**) were taken year-round using automated water samplers (similar to Isco 3700R automated refrigerated water sampler, Teledyne Isco, Lincoln, NE, USA) installed aboard commercial vessels (GTS Finnjet, m⁄s Finnpartner, m⁄s Finnmaid) travelling between Finland and Germany (Fig. 1). The ship-of-opportunity sampling locations varied within the study period: during 1993–1995, sampling took place at specific times, and from 1996 onward, sampling took place at given longitudinal positions (see Rantajärvi 2003 for a more detailed description of the sampling method).

Rantajärvi et al. (1998) argue that the single ship-of-opportunity-sampled depth may represent the productive layer in the Baltic Sea, but we (Majaneva et al. 2009) found that the ship-of-opportunity samples underestimated the abundance of blooming cyanobacteria during a strong stratification. The argument of Rantajärvi et al. (1998) may hold when there is no stratification or it is weaker than in summer. The ship-of-opportunity samples were kept in darkness in a refrigerator until they were brought into the laboratory, where they were further sampled for DNA and preserved with acid Lugol's solution (Willén 1962). Time between sampling and preservation was approximately 36 h, giving time for additional 15 % heterotrophic growth of *Prymnesium* species in the ship-of-opportunity samples (Granéli et al. 2012). The integrated samples were preserved with acid Lugol's solution immediately after sampling. Thus, the two types of monitoring samples may represent slightly differing communities.

3.2.2. Sea-ice and water samples (III-V)

Sea-ice samples were obtained from eight locations: four fast-ice stations and four drift-ice stations (Fig. 1). The fast-ice stations were along the coast of southern Finland; two locations near the Tvärminne Zoological Station, in March 2001 (**III**) and February–March 2006 (**IV**), one at Vuosaari harbor in Helsinki, in March 2007 (**IV**) and one near the town of Kotka, in March 2010 (**V**).

The Bothnian Sea and Bothnian Bay (**IV**) drift-ice stations were sampled during the cruise of R/V Maria S. Merian, in March 2006, while the Gulf of Finland drift- and pack-ice stations (**V**) were sampled during R/V Aranda sea ice cruise, in March 2010.

Samples were obtained using three different approaches. 1) Most of the samples (**III-V**) were taken using a motorized CRREL-type ice-coring auger (9 cm internal diameter; Kovacs Enterprises, Indianapolis, IN, USA). 2) The pancake-ice sample from the Bothnian Sea (**IV**) was obtained using a crane and a metal basket onboard the R/V Maria S. Merian. 3) The samples from Vuosaari harbour (**IV**) were collected using a hand ice saw. In addition, three replicate slush samples were taken from a 50 cm × 50 cm square with a hand shovel from Kotka (**V**).

The ice cores were either cut into pieces (**III, V**) or treated as one sample (**IV**). At each station during the R/V Aranda cruise (**V**), five ice cores were taken and sectioned into five pieces: surface, upper intermediate, middle, lower intermediate and bottom sections. In order to get all measurements from the same samples, we put all five surface sections into one plastic bag, all five bottom section into another plastic bag etc. and crushed them inside the bags and left to melt in a bucket in darkness at +4°C without filtered seawater. Slush (**V**) was put to melt in a basket in darkness at +4°C. The Vuosaari harbour samples (**IV**) were melted likewise without filtered seawater. In contrast, the earlier samples (**III-IV**) were left to melt in darkness while submerged in 0.2-µm-filtered seawater (salinity 6) at +4°C. The latter method was recommended by Garrison & Buck (1986) and Kottmeier & Sullivan (1988) because of a loss of cells in samples melted without seawater. However, Kaartokallio (2004) and our unpublished data (Rintala et al. unpublished) suggest that melting in filtered seawater is not necessary for the Baltic Sea ice samples. The reason for this is probably the lower salinity of the brine in the Baltic Sea ice and hence reduced osmotic stress during melting than in oceanic ice. The ice and slush samples were subsampled for DNA extraction and cell fixation after melting.

The water samples (**IV**) were collected from 2 m depth using a Limnos water sampler (Limnos Ltd., Turku, Finland) in Storfjärden, Tvärminne Archipelago, in February 2006. Under-ice water (**V**) was sampled submersing one-liter bottles in the corer holes during the R/V Aranda cruise.

3.3. Light microscopy (I-III, V)

The morphologies of the cultured strains (**I, III**) were examined using Leitz Aristoplan light microscope (Ernst Leitz Wetzlar GmbH, Wetzlar, Germany) or Leitz DM IRB inverted microscope with attached Leica DC300F (Leica Microsystems, Wetzlar, Germany) or Polaroid DMC 1 (Polaroid Corporation, Cambridge, MA, USA) digital camera and Leica DC Twain v.4.1.5.0 image acquisition software. Cell dimensions were measured from light micrographs of alive (**I**), Lugol's-fixed (**I, III**), or glutaraldehyde-fixed and Calcofluor-stained (Fritz & Triemer 1985, **III**) cells.

The monitoring samples (**II-III**) were counted in accordance with HELCOM monitoring guidelines (HELCOM 2008), using inverted light microscopes similar to the Leitz DM IRB with phase contrast. Subsamples of 10, 20, 25 or 50 ml were settled for at least 19 h, and several transects covering 1.2–5 % of the cuvette bottom or 20–200 randomly chosen fields of view and 400–600× magnifications were used for cell enumeration of the Prymnesiales species

(**II**). *Heterocapsa arctica* subsp. *frigida* (**III**) abundance was estimated on a five-rank abundance scale (1 = very sparse, 2 = sparse, 3 = scattered, 4 = abundant, 5 = dominant) from 50 ml of at least 24-h-settled subsamples. As the focus of the studies was more on the occurrence of Prymnesiales species (**II**) and *Heterocapsa arctica* subsp. *frigida* (**III**) and not on the absolute numbers, the HELCOM (2008) method was considered sufficient.

The cell enumeration of sea-ice organisms (**V**) was made with a Leica DMIL light microscope from glutaraldehyde-fixed samples where acid Lugol's solution was added prior the counting. 50 ml subsamples were settled for 24 h (Utermöhl 1958). Larger organisms were counted with 10×/12.5 objectives from the whole cuvette bottom, while smaller organisms were counted with 25× or 40×/12.5 objectives from 60 or 120 randomly chosen fields of view (depending on the density of the organisms) distributed evenly over the cuvette bottom.

3.4. Electron microscopy (I-III, V)

Cryptomonads are easily identified by a distinctive wobbling swimming pattern, asymmetric cell shape and the presence of an anterior depression which extends inside the cell. Taxonomy below the class level is difficult because of low number of characters visible under the light microscope. For more detailed analysis of cell surface and ultrastructure, *Rhinomonas nottbecki* cells (**I**) were fixed for 90 min in 2 % glutaraldehyde buffered to pH 7.2 with 100 mM sodium cacodylate and osmotically balanced with 800 mM sucrose. The cells were transferred into a buffer containing 100 mM sodium cacodylate and 800 mM sucrose and kept at +4 °C, until postfixed with 1 % OsO_4 at room temperature for 1 h.

After postfixation, cells were dehydrated in series of ethanol (70 %, 96 % and 100 %) and critical point dried, using a Bal-Tec CPD 030 Critical point drying unit (Bal-Tec Union Ltd, Liechtenstein). The material for cell exterior examination was mounted with carbon tape on aluminum stubs and coated with 5 nm of colloid platinum using a Quorum Q150T S sputter (Quorum Technologies Inc., Guelph, Canada) before visualization with a FEI Quanta FEG 250 (FEI, Hillsboro, Oregon, USA) scanning electron microscope (SEM). The material for cell interior examination was gradually embedded in Lowicryl HM20 or Epon (TAAB Laboratories Equipment Ltd, Berkshire, UK) and thin sectioned using a Leica EM Ultracut UC6i (Leica Mikrosysteme GmbH, Vienna, Austria) ultramicrotome. The thin sections were collected on copper grids and examined with a JEOL JEM-1200EX (JEOL Ltd, Tokyo, Japan) transmission electron microscope (TEM), using 60–90 kV tension.

To examine the internal structures of *Rhinomonas nottbecki* cells (**I**) in 3D, the cells were prepared for Serial Block-Face Scanning Electron Microscopy (SBF-SEM) (Denk & Horstmann 2004). The protocol is based on Deerinck et al. (2010) and includes ferrocyanide reduced osmium tetroxide postfixation (2 % OsO_4), thiocarbohydrazide-osmium liganding and subsequent uranyl acetate (1 % UA) and *en bloc* lead aspartate staining. After staining, the cells were dehydrated and embedded in Durcupan ACM (Fluka, Sigma-Aldrich). The Durcupan was mixed based on EMS recommendations, where the parts A, B, and D are mixed first and the part C is added only at the very last step of the embedding process (http://www.emsdiasum.com/microscopy/technical/datasheet/14040.aspx). The blocks were

imaged with a FEI Quanta FEG 250 using a backscattered electron detector (Gatan Inc., Pleasanton, CA, USA). Imaging parameters were 2.5 kV beam voltage, 3.0 spot size, 0.3-Torr pressure with XY resolution of 14 nm/pixel. The microscope was equipped with a microtome (3View; Gatan), which allowed serial imaging of block faces with increments of 30 nm. The images were processed using Microscopy Image Browser, a program developed by Ilya Belevich and written under Matlab (Mathworks, Natick, MA, USA) environment.

Glutaraldehyde-fixed *Heterocapsa arctica* subsp. *frigida* cells (**III**) were treated according to Hansen (1995b) and examined using a Zeiss DSM 962 SEM (Carl Zeiss, Oberkochen, Germany). Shadow-cast whole mounts were prepared according to Iwataki et al. (2003), and the scales were visualized with a JEOL JEM-1010 TEM using 90 kV acceleration voltage. Another part of the fixed sample was processed according to Jensen & Moestrup (1999). A Leica ultramicrotome was used for thin sectioning. The thin sections were collected on plastic-covered copper grids and examined with a JEOL JEM-1200EX TEM using 60 kV tension for cell ultrastructure characters.

To visualize the extracellular body scales of Lugol's-fixed Prymnesiales (**II**), whole-mounts were prepared as explained by Moestrup & Thomsen (1980). Ten drops from each sample were collected with a glass Pasteur pipette from the bottom of the LM cell enumeration cuvette or from the bottom of the 300 ml stored sample. The drops were placed in an electron microscope grid covered by a film of formvar and carbon. The grids were kept in a fume hood until dry. The salt crystals formed were removed by submerging each grid three times in sterile double-distilled water, letting the grids air-dry after each submergence. Five grids were stained with uranyl acetate (UA) by placing them for 20 min in 2 % neutralized UA followed by three rinses in double-distilled water to remove the excess stain. These grids were examined at 60 kV with a JEOL 1200 EX TEM. Five grids were placed at a 20° angle inside a shadow-casting sputter JEOL JEE 4B and coated with a 40:60 mixture of gold:palladium and examined using a 72-kV acceleration voltage in a JEOL 1011 TEM.

3.5. Extraction of nucleic acids (I-V)

3.5.1. DNA extraction (I-V)

DNA was extracted using the phenol-chloroform method (Maggs & Ward, 1996, **I-V**) or DNeasy® Plant Mini Kit (Qiagen GmbH, Hilden, Germany) (**I**). The phenol-chloroform method results in long DNA strands ideal for cloning while the commercial kits usually result in more fragmented DNA. The DNeasy kit worked much better with the cryptomonad cells (**I**) and hence, was preferred.

Prior to DNA extraction the cells in culture (**I, III**) were collected with low centrifugation to the bottom of a microcentrifuge tube and DNA lysis buffer was added. The cells in environmental samples were trapped onto membrane filters (**II, IV-V**). The material was size fractionated so that DNA was extracted from 2.0–20 μm (**II**), 0.65–20 μm (**IV**) or 0.2–20 μm (**V**) fraction. The filters were dipped into liquid nitrogen and stored at -70 °C and soaked in the DNA lysis buffer prior to DNA extraction (**II, V**), or just soaked in the DNA lysis buffer and kept cool until DNA extraction (**IV**). After extraction, DNA was further purified using a High Pure PCR Template Preparation Kit (Roche Diagnostics, F. Hoffmann-La Roche

AG, Basel, Switzerland) according to the manufacturer's instructions (**IV**).

3.5.2. RNA extraction and cDNA synthesis (IV)

The cells for total RNA extraction were trapped onto 0.2-µm-pore-size filters (size faction 0.2–40 µm). The filters were dipped into liquid nitrogen prior to extraction with the Qiagen RNeasy Plant Mini Kit. Complementary DNA (cDNA) was synthesized from rRNA, using SuperScript™ III Reverse Transcriptase (Invitrogen Corp. (now Life Technologies Corp.), Carlsbad, CA, USA) according to the manufacturer's instructions with the primers UNI7F and UNI1534R (Moon-van der Staay et al. 2001).

3.6. DNA amplification, cloning and sequencing (I-V)

3.6.1. Polymerase chain reaction (I-V)

Polymerase chain reaction (PCR) was used to amplify 18S rRNA genes (**I-II, IV-V**) and ITS region (**I, III**). The amplification was done with published or with primers designed for this study, using polymerases provided by several manufacturers (Table 2). The denaturation and annealing temperature and time, extension time and number of cycles varied depending on the length of the amplified region and the polymerase. A typical PCR cycle using Taq DNA polymerase (ABgene, Epsom, UK) was initial denaturation of 3 min at 94 °C, followed by 30 cycles of 1 min at 94 °C, 1 min at 50 °C, 2 min at 72 °C, and a final extension of 5 min at 72 °C (**III**). The PCR products were purified with an Illustra GFX™ PCR, DNA and Gel Band Purification Kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK), with a Montage SEQ 96 Sequencing Reaction Cleanup Kit (Merck Millipore, Billerica, MA, USA) or with a High Pure PCR Product Purification Kit (Roche Diagnostics) according to the manufacturers' instructions.

3.6.2. Cloning (I-IV)

In order to differentiate between multiple copies of the ITS-region (**I, III**) or several taxa in one sample (**II, IV**), PCR reactions were cloned with the pGEM Easy cloning kit (Promega, Madison, WI, USA) (**I, III**), PCR Cloning$_{plus}$ kit (Qiagen, Hilden, Germany) (**II**) or the TOPO® Cloning Kit for Sequencing (Invitrogen Corp. (now Life technologies Corp.), Carlsbad, CA, USA) (**IV**). Positive colonies were picked with a toothpick and

Table 2. The polymerases used in the thesis.

Polymerase	Used in paper
Taq DNA polymerase (ABgene, Advanced Biotechnologies Ltd., Epsom, UK)	I, II, III, IV
DyNAzyme™ II DNA Polymerase (Finnzymes Oy, Vantaa, Finland)	I
FastStart Taq DNA Polymerase (Roche Diagnostics, Basel, Switzerland)	I
KOD -Plus- ver. 2 (TOYOBO Co. Inc, Osaka, Japan)	V

dipped into the PCR reaction mixture of Taq DNA polymerase. The amplification was done using the primers M13F (5' GTA AAA CGA CGG CCA G 3') and M13R-pUC (5'-CAG GAA ACA GCT ATG AC-3') (**I-II, IV**) or the primers T7 (5'-TAA TAC GAC TCA CTA TAG GG-3') and SP6 (5'-ATT TAG GTG ACA CTA TAG AA-3') (**III**). The reactions were then purified with the Montage SEQ 96 Sequencing Reaction Cleanup Kit (**I, IV**), the USB ExoSAP-IT® (Affymetrix, Santa Clara, CA, USA) (**II**) or the Illustra GFX™ PCR, DNA and Gel Band Purification Kit (**III**) according to the manufacturer's instructions.

3.6.3. Sequencing (I-V)

Cycle sequencing of the PCR products (**I**) or positive clones (**I-IV**) was carried out using the Big Dye™ terminator (Applied Biosystems, Foster City, CA, USA) cycling conditions. The samples were loaded on an automated sequencer 3730xl (Applied Biosystems) at Macrogen Inc, Seoul, Korea. The PCR products of the paper **V** samples were mixed equally and a DNA library was prepared by use of a GS FLX Titanium Rapid Library Preparation Kit (Roche) according to the manufacturer's instructions. This library was then amplified on beads by emulsion polymerase chain reaction, and the amplified fragments in the DNA library were pyrosequenced on a 1/4 section of picotiterplate using 454 GS FLX Titanium system and reagents (Roche) in Japan.

3.6.4. Sequence quality (I-V)

The Sanger sequence chromatograms (**I-IV**) were manually checked with the program Chromas Lite 2.1 (Technelysium Pty Ltd) to ensure correct base-calling, after which chimeric sequences were identified using the Chimera Detection (Cole et al. 2003) and Bellerophon (Huber et al. 2004) programs (**IV**).

The 454 sequences (**V**) were processed, using mothur v.1.29.0 (Schloss et al. 2009) and following the Schloss SOP pipeline (Schloss et al. 2011) in www.mothur.org/wiki/ (accessed January 21, 2013). To ensure good quality of the sequences, the trim.seqs command was used to cut sequences when the average quality score over a 50-bp window dropped below 25, and to eliminate sequences with >6 homopolymers, sequences with ambiguous bases, and sequences with >zero mismatch in the barcode and the primer sequence. The unique sequences were aligned against the SILVA reference file provided in the mothur-wiki pages. The alignment was filtered so that all sequences overlapped in the same region. The pre.cluster command was used to merge sequences that were within 2 bp of a more abundant sequence. Chimeric sequences were identified using Uchime (Edgar et al. 2011) and removed. Schloss et al. (2011) report that following the SOP pipeline the overall chimera rate decreases to 1 %. There were 242 049 on average 230-bp-long (including the V8 and V9 variable regions of the 18S rRNA gene) sequences left after the pre-processing.

3.7. Phylogenetic analyses (I-IV)

3.7.1. Alignment of sequences (I-IV)

DNA amplification, cloning and sequencing was conducted in order to make phylogenetic analyses based on the DNA sequences. The good quality sequences were aligned using the programs mothur (Schloss et al. 2009) (**I**) and MAFFT (Katoh & Toh 2008) with

Q-INS-i (**I, II**) or FFT-NS-2 (**IV**) strategy and gap-opening penalty of 1.53 (**I-IV**) and gap-extension penalty of 0.123 (**I-II, IV**) or 0 (**III**).

The Pyrenomonadaceae nucleus- and nucleomorph-encoded 18S rRNA gene sequences (**I**) were aligned with all cryptophyte nucleus and nucleomorph sequences, respectively, from the SILVA database (Quast et al. 2013) as a reference. The Prymnesiales 18S rRNA gene sequences (**II**) were aligned with the Prymnesiales sequences published in GenBank and the supplementary material SSU FIN.txt of Edvardsen et al. (2011). The ITS region of *Heterocapsa arctica* subsp. *frigida* (**III**) was aligned with published *Heterocapsa* ITS-region sequences. All alignments were visually checked and adjusted.

3.7.2. Generating the phylogenetic trees (I-IV)

Maximum likelihood trees and bootstrap support values were calculated using GARLI (Zwickl 2006) (**I-II, IV**) or Phylip (Felsenstein 2004) (**III**). Proper evolutionary models were selected with AICc criterion in jModelTest (Posada 2008). Neighbor-joining analyses were performed using Phylip (**III**) and maximum parsimony analyses using TNT (Goloboff et al. 2008) (**II-III**). Posterior probabilities were calculated with MrBayes (Ronquist et al. 2012) (**I-II, IV, this thesis**).

The trees were rooted using outgroups. By definition, an outgroup is a group of organisms that is closely related to the group of interest but less closely than those inside the group of interest. This assumes that during the course of evolution, the outgroup diverged first from the ancestral group before the group of interest. The Pyrenomonadaceae nuclear 18S rDNA data set included all kathablepharids as an outgroup while three rhodophytes were used as an outgroup in the nucleomorph data set (**I**). The Prymnesiales had four *Isochrysis* sequences as an outgroup (**II**), while the *Heterocapsa* tree was rooted with *H. illdefina* (**III**).

The Syndiniales and Fungi trees were generated with a slightly different approach (**IV**). Since those groups include several hundreds and thousands of sequences, our sequences were assigned in the respective groups using the BLAST network service (Altschul et al. 1997) and their position was visualized in phylogenetic trees using the closest hits and some randomly chosen representatives of the respective groups.

3.8. Richness and diversity analyses (IV-V)

3.8.1. Operational taxonomical units (OTUs) (IV-V)

Precise taxonomic designation is not possible with methods that utilize only 18S rRNA gene sequences, and sequences are used as a proxy for species or other taxa. As discussed above, groups of similar sequences are referred to as operational taxonomic units (OTUs). The sequences were clustered into OTUs at 97 % similarity level, using mothur (**IV-V**). We chose the 97 % similarity level for our analyses for three reasons. Firstly, pyrosequencing (**V**) is associated with a high error rate which is mainly caused by miscounted homopolymeric runs that occur in otherwise high quality regions of the sequences (Kunin et al. 2010). These errors may be abundant and produce spurious OTUs, and therefore, Kunin et al. (2010) recommended using the 97 % similarity level to lower the amount of those OTUs. Secondly, Bachy et al. (2013) showed that the multiple

alignments that are needed for assigning the pyrosequences into OTUs in mothur may include small errors which may lead to a 10- or even 100-fold overestimation of OTUs at high similarity level (99 %). In our dataset (**V**), the number of 99 % OTUs was double to the number of 97 % OTUs. Thirdly, the variability in the 18S rDNA is not equal within different eukaryotic lineages (Caron et al. 2009), and there is no universal sequence similarity level that could be used. Within lineages evolutionary rates may be estimated and appropriate similarity levels are needed for each group. When comparisons are made across all eukaryotes, as here (**IV-V**), an arbitrary similarity grouping is required. We found earlier (Suutari et al. 2010) that the 97 % similarity level is conservative enough to exclude errors but sensitive enough to discriminate the lineages.

Taxonomic assignment of the 97 % OTUs (**IV**) was generated using the BLAST network service (Altschul et al. 1997). In paper **V**, SILVA database release 104 (Quest et al. 2013) within Qiime program package (Caporaso et al. 2010) was used. If SILVA failed to find an assignment, the BLAST network service was used. If no taxonomic assignment was found, the OTU was classified as Eukaryota.

3.8.2. OTU richness and diversity estimates (IV-V)

Based on the abundance and incidence of OTUs and on sequence divergence, three types of richness and diversity estimates were calculated. The abundance-based Chao1 (**IV**) and Shannon and Simpson's indices (**IV-V**) were calculated using mothur. The incidence-based Chao2 and ICE (**IV**) were calculated using EstimateS 8.2.0 (Colwell 2005). The divergence-based PD was calculated with the program Phylocom 4.1 (Webb et al. 2008) from 20 replicate maximum likelihood tree searches that were calculated using GARLI version 1.0.659 (Zwickl 2006) with K80+G model (**IV**) or with mothur (**V**). The other divergence-based index θ (**IV**) was calculated with the program Arlequin 3.5 (Excoffier et al. 2005).

The shared OTUs among groups of samples were visualized using Venn diagrams that were calculated using mothur (**IV-V**). To examine the difference of the community composition structure of the samples, unweighted UniFrac measures (**IV**) were calculated from a neighbor-joining tree constructed in Phylip (Felsenstein 2004), using mothur. Analysis of molecular variance (AMOVA), calculated both with the membership- and the structure-based Chi-square distance using mothur, was used to compare the differences in the community membership and structure of the samples (**V**).

4. RESULTS AND DISCUSSION

4.1. Taxonomic studies of some Baltic Sea flagellates (I-III, this thesis)

The first part of my thesis includes three taxonomic studies of flagellates living in the Baltic Sea. A cryptomonad (**I**) and a dinoflagellate (**III**) strain were collected from the Hanko Peninsula area. Both strains had characters that showed them to be previously undescribed taxa. During the autumn and winter of 2007–2008, a haptophyte dominated the Baltic Sea water community (**II**). The blooming species had characters not previously reported in nature. The results of these three studies are presented here as is the discussion of the taxonomic status of the dinoflagellate.

4.1.1. Taxonomy of Rhinomonas nottbecki (Cryptomonadales) and the phylogeny of the family Pyrenomonadaceae (I)

Rhinomonas nottbecki sp. inedit., referred as *Rhinomonas nottbecki*, will be published independently later, and the occurrence of the name in this thesis (the thesis summary and the paper **I**) is for convenience only and has no standing in nomenclature.

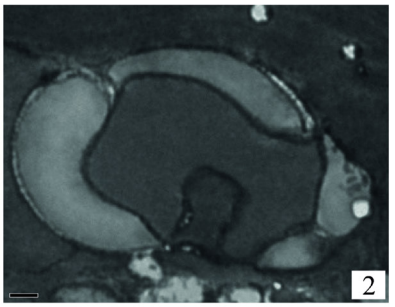
The synapomorphy of cryptomonads belonging to the family Pyrenomonadaceae is the nucleomorph that bisects the pyrenoid (Clay et al. 1999). This is a characteristic of the species we observed (Fig. 2).

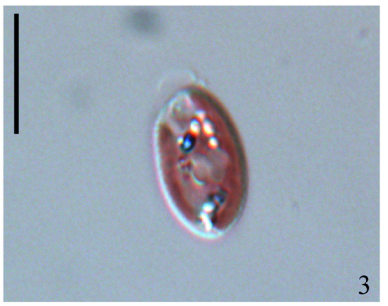
Rhinomonas nottbecki (**I**) can be distinguished from the other Pyrenomonadaceae species using plate morphology, cell size, and the location of vestibulum. A synapomorphy of the genus *Rhinomonas* is hexagonal plates, while the genera *Rhodomonas* and *Pyrenomonas* have rectangular plates and the genus *Storeatula* has no plates (Hill & Wetherbee 1988, 1989, Hill 1991). In addition, the *Rhodomonas* species have a furrow, absent from the species observed here.

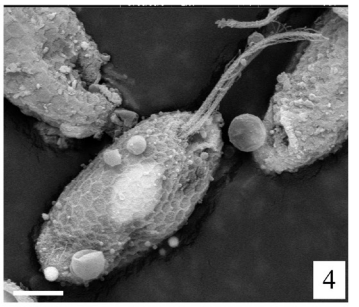
Within the genus *Rhinomonas*, *R. nottbecki* with a length of 10–17 µm (Fig. 3) is longer than *R. pauca* (7–9 µm) *R. fulva* (5–7 µm), *R. lateralis* (5–7 µm) and *R. fragarioides* (5.5–6 µm), but overlaps with that of *R. reticulata* (8–14 µm) (Hill & Wetherbee 1988, Novarino 1991a, 1991b). The vestibula of the smaller *Rhinomonas* species opens about one-third to one-fourth cell length away from apex, and this differs from the one-fifth in *R. reticulata* and one-sixth in *R. nottbecki* (Figs. 3 and 4). The result is that the appearance of *R. reticulata* and *R. nottbecki* is less rhinote (the cell having not so clear 'nose' when viewed from the side). The *R. nottbecki* plates are 0.4 µm long (Fig. 4), half the size of *R. pauca* plates.

To distinguish *R. nottbecki* and different varieties of *R. reticulata*, plate size and cell compression (mean cell thickness to mean cell width ratio) is used. The plates of *R. nottbecki* are in the same range than in *R. reticulata* var. *reticulata* and *R. reticulata* var. *atrorosea* (Novarino 1991a, 1991b). *R. reticulata* var. *compressa* and *R. reticulata* var. *eleniana* have smaller plates. *R. nottbecki* is more compressed (ratio 1.13, n = 23) than *R. reticulata* var. *reticulata* (1.04) and *R. reticulata* var. *atrorosea* (1.09) (Novarino 1991a, 1991b).

The DNA-analysis shows further evidence to distinguish the species. The nucleus-







Figs. 2-4. *Rhinomonas nottbecki* micrographs. – **Fig. 2.** Transmission electron micrograph showing the pyrenoid covered with starch sheets and the invaginated nucleomorph. Scale bar 0.2 µm. – **Fig. 3.** Light micrograph of live cell. Scale bar 10 µm. – **Fig. 4.** Scanning electron micrograph showing the hexagonal plates. Scale bar 2 µm.

encoded 18S rRNA gene of *R. nottbecki* differed from other Pyrenomonadaceae sequences by 0.07–2.83 % which is equivalent to 1–39 bases difference. At site 583 within the variable region V4 of the nucleus 18S rRNA gene of *Rhodomonas baltica* (AB241128), *R. nottbecki* had the base C while the other Pyrenomonadaceae had the base A. This was the only difference between *R. nottbecki* and *R. reticulata* in the nuclear 18S rRNA gene. Although the family Pyrenomonadaceae formed a well-supported clade (1.0/100 posterior probability/maximum likelihood bootstrap support values) in the nucleus-encoded 18S rRNA gene tree of cryptomonads, *Goniomonas* and kathablepharids, the species' phylogenetic relations remained unresolved within the family.

Based on our results and the results from studies of Hoef-Emden and colleagues (Hoef-Emden et al. 2002, Hoef-Emden & Melkonian 2003, Hoef-Emden 2007) it can be concluded that the 18S rRNA gene is too conservative within cryptomonads for reliable species discrimination. Therefore, the 18S rRNA gene of the reduced endosymbiotic red algal nucleus (nucleomorph) and the more variable ITS region of the nuclear ribosomal operon were amplified following Hoef-Emden and colleagues (Hoef-Emden et al. 2002, Hoef-Emden & Melkonian 2003, Hoef-Emden 2007).

The *R. nottbecki* nucleomorph sequence differed by 2.07–10.01 % (31–133 bases) from other Pyrenomonadaceae sequences. The differences were within the sites 463–474 (V2), 841–978 (V4) and 1739–1748 (V8). The family Pyrenomonadaceae sequences formed a well-supported clade (1.0/77), and *R. nottbecki* strains formed a clade of their own (1.0/100) in the nucleomorph-encoded 18S rRNA tree (Fig. 5).

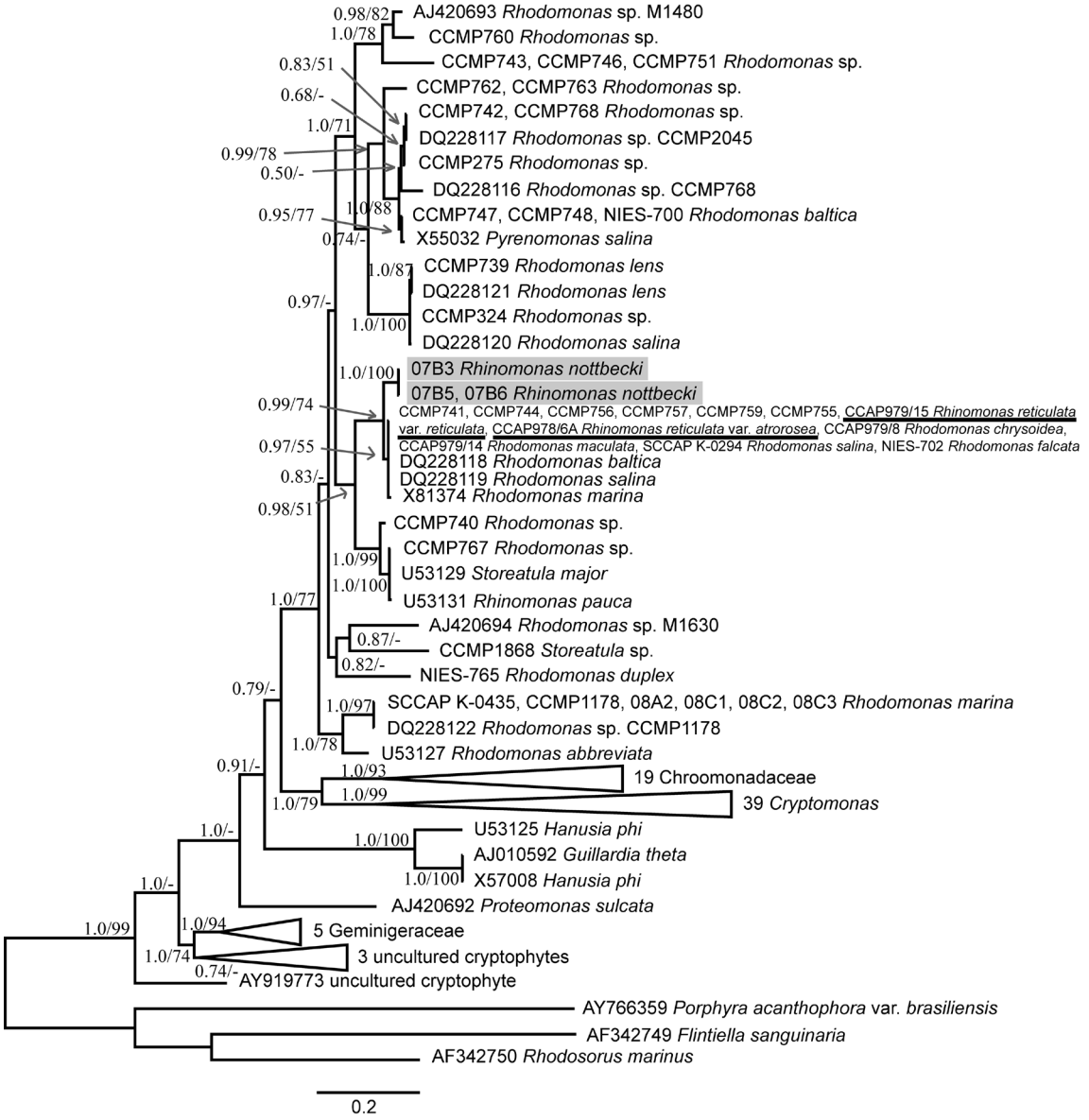


Fig. 5. Consensus Bayesian tree based on nucleomorph-encoded 18S rRNA gene sequences of members of cryptomonads. The sequences are denoted with the strain names (sequenced within this study) or accession numbers (derived from GenBank). The sequences of *Rhinomonas nottbecki* n. sp. strains are highlighted with grey background and *Rhinomonas reticulata* sequences are underlined. Posterior probabilities (left) and maximum likelihood bootstrap support values (right) are shown near the internal nodes. Three rhodophyte species were used as an outgroup and their support values are not shown. Posterior probabilities were calculated across the GTR model space while the maximum likelihood bootstrap values were calculated with the TVM+G model.

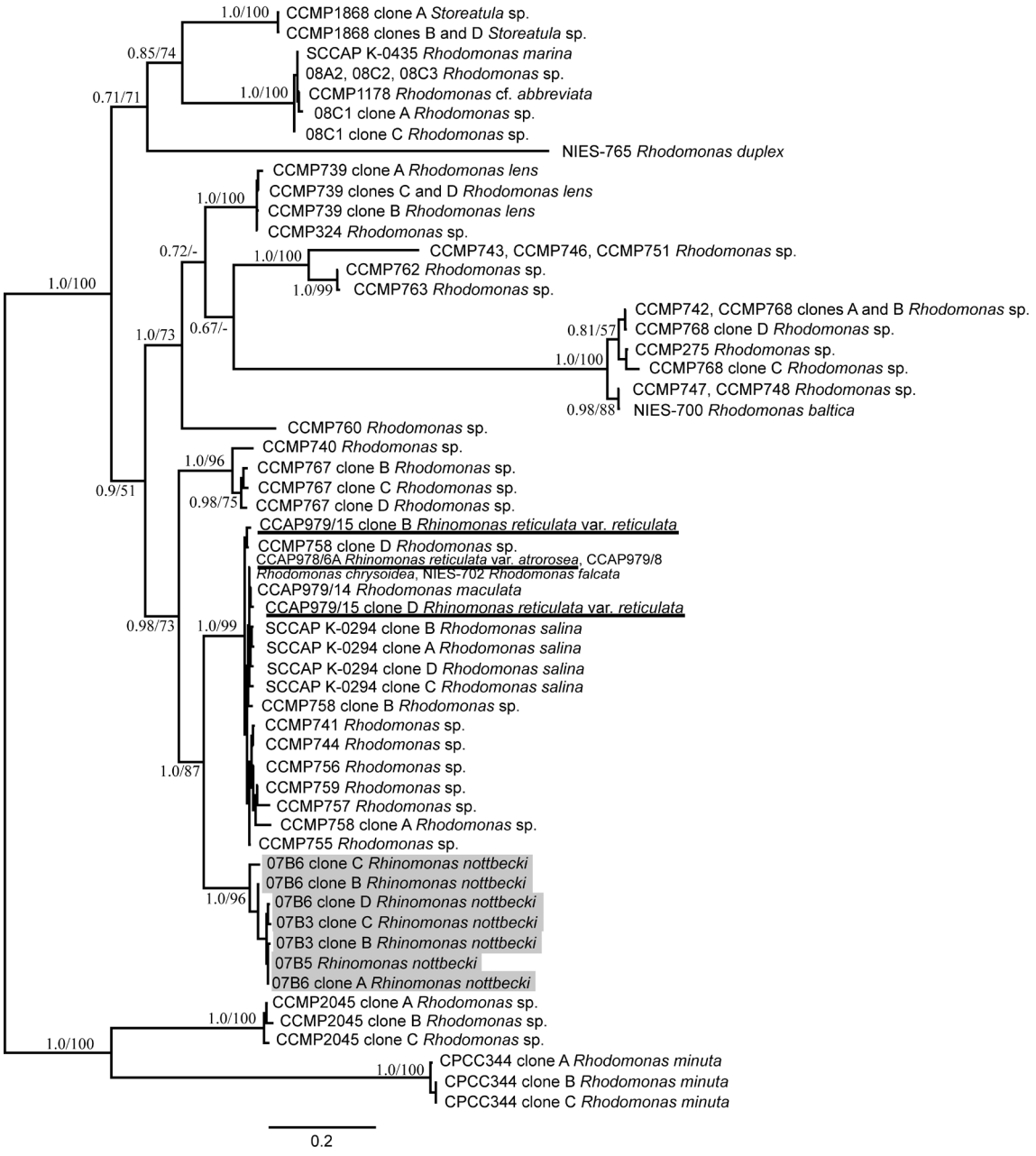


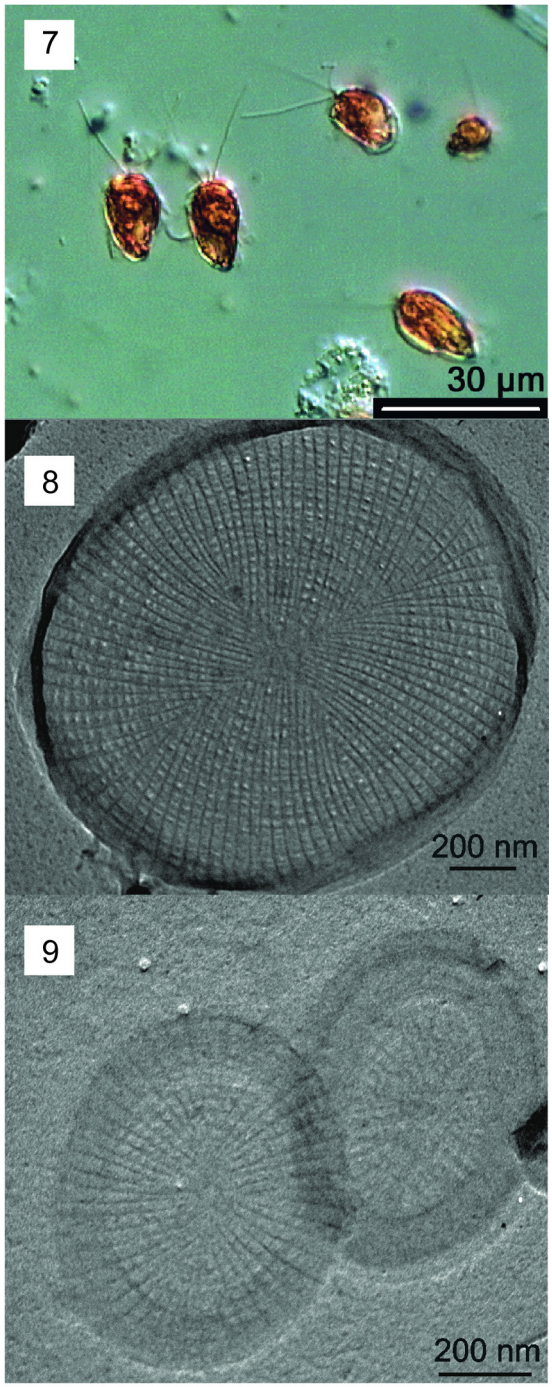
Fig. 6. Consensus Bayesian tree based on ITS region data of the members of the family Pyrenomonadaceae. The sequences are denoted with the strain names. The sequences of *Rhinomonas nottbecki* strains are highlighted with grey background and *Rhinomonas reticulata* sequences are underlined. Posterior probabilities (left) and maximum likelihood bootstrap support values (right) are shown near the internal nodes. Posterior probabilities were calculated across the GTR model space while the maximum likelihood bootstrap values were calculated with TIM3+G model.

The intraspecific difference of the ITS region of *R. nottbecki* was 0–4.99 % (0–70 bases), including large indels in the ITS1 and ITS2 regions. The amplified region of *R. nottbecki* differed by 11.25–31.35 % (> 118 bases) from other Pyrenomonadaceae sequences. *R. nottbecki* strains formed a well-supported clade (1.0/96) in the ITS region tree (Fig. 6). Consequently, the evidence from our morphological and molecular analyses permitted us to describe *R. nottbecki* as a new species belonging to the family Pyrenomonadaceae.

4.1.2. Wintertime blooming Prymnesium polylepis (II)

In late October 2007, a considerable increase in the number of haptophytes was observed in several parts of the Baltic Sea within routine national monitoring programmes. Based on cell size and shape, and the lengths of the flagella and haptonema under the light microscope (Fig. 7), the cells were identified as members of the order Prymnesiales.

Most of the cells were 10–20 µm long, agreeing with *Chrysochromulina birgeri* and *C. limonia* found in the Baltic Sea. *C. birgeri* is a cold-water species that has large scales with hornlike projections and frequently four flagella (Hällfors & Niemi 1974, Hällfors & Thomsen 1979), which were not found in



Figs. 7-9. The alternate stage *Prymnesium polylepis* micrographs. – **Fig. 7.** Light micrograph of Lugol's fixed cells. Scale bar 30 µm. – **Fig. 8.** Shadow-casted transmission electron micrograph of the large scale type of the alternate stage *Prymnesium polylepis*. Scale bar 0.2 µm. – **Fig. 9.** Shadow-casted transmission electron micrograph of the small scale type of the alternate stage *Prymnesium polylepis*, showing distal and proximal sides. Scale bar 0.2 µm.

the blooming species. Similarly, *C. limonia* (Jensen 1998) has easily distinguishable scales with spines.

Transmission electron microscopic examination of the scales was needed to confirm the identity of the blooming species. Two differently-sized flat scales belonging to one species were found in samples from October 2007 to April 2008. Sample BMPH12 on May 7 2008 included also scales of *Haptolina ericina*. Sample SB5 on May 27 2008 included only scales of *H. ericina* and scales that resembled those of *H. fragaria* (Jensen 1998).

In the species with two types of scales, the smaller scales were on average 0.76 × 0.58 µm, and the larger 1.50 × 1.14 µm (Figs. 8 and 9). These scales were very similar to those of *Chrysochromulina mantoniae* and the alternate stage of *Prymnesium polylepis* (Jensen 1998). However, we found neither the stout spine scales typical of *C. mantoniae* nor the spine-bearing scales typical of alternate stage *P. polylepis*.

For that reason, we sequenced the 18S rRNA gene from two samples (stations SB5 and SB6, Fig. 1) taken during the peak of the bloom in March 2008 to identify the species. Edvardsen et al. (2011) showed that species of Prymnesiophyceae can be distinguished based on the 18S rRNA. There is no *C. mantoniae* reference sequence in GenBank, but we are confident that our sequences were not *C. mantoniae*, since all sequences clustered with named GenBank sequences and not as a new, distinct branch (Fig. 10). The grouping of our isolates with *P. polylepis* was well supported (98/96/1.0 maximum likelihood/maximum parsimony/posterior probability). The intraspecific difference of the 18S rRNA gene of *P. polylepis* sequences was 0–0.49 %, equivalent to up to 7 bases difference. The sequence divergence between our sequences and other species of the genus *Prymnesium* in GenBank was 1.04–3.69 %.

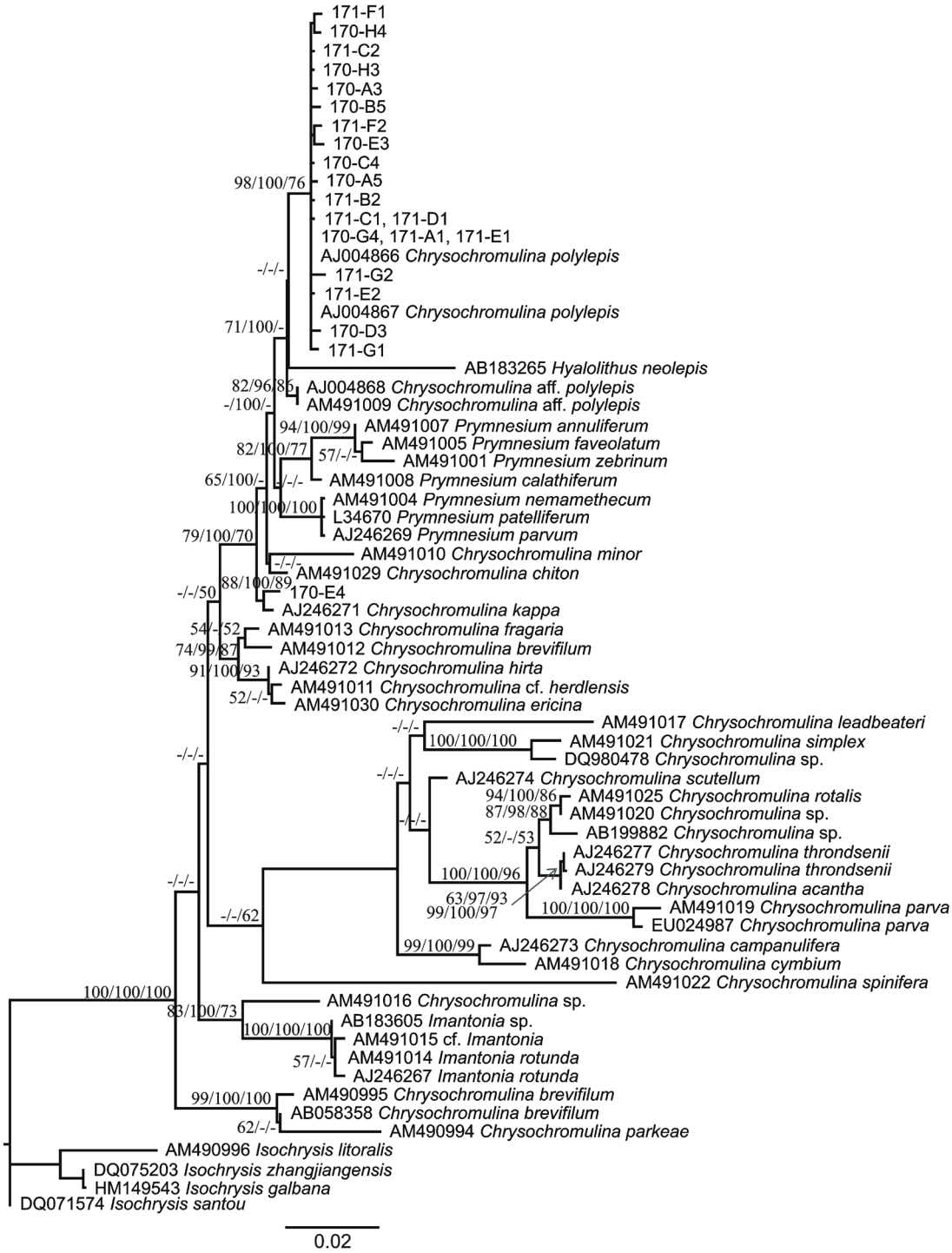


Fig. 10. Maximum likelihood tree based on the 18S rRNA gene sequence data of Prymnesiales samples sequenced in this study (samples SB5 = 170 and SB6 = 171) and received from GenBank (accession numbers shown). The numbers at the nodes of the tree represent the maximum likelihood bootstrap (> 50), posterior probability (> 95) and maximum parsimony bootstrap (> 50) values.

Our molecular evidence combined with the flat scale morphology led to the conclusion that the material collected was alternate stage *P. polylepis*, although the cell size was larger and scale size smaller than in the original strain described by Paasche et al. (1990).

Prymnesium polylepis dominated the phytoplankton community from December 2007 to May 2008, accounting for up to 80–88 % of the total phytoplankton biovolume, and reaching maximum abundances in April–May (Fig. 3 in **II**). The onset of the bloom coincided with a significantly lower wind and precipitation in the whole Baltic proper area (Fig. 4 in **II**) and a strong thermocline at 30–40 m depth (Yhlen & Andersson 2007, Thorstensson & Yhlen 2007). The following winter was the mildest recorded in south-eastern Sweden since 1858–1859 (Hellström 2008) with 2–4°C higher than normal water temperatures in the Baltic Sea (Grafström 2008, Lake & Grafström 2008). There was no stratification in the uppermost 30–50 m in January–March 2008 (Hansson & Andersson 2008).

The calm and sunny weather in October resulted in high light availability for the season and low turbulence above the thermocline. This made it possible for *P. polylepis* to build up a considerable biomass which persisted throughout the winter. Possible allelopathic effects that may have inhibited the growth of grazers and competitors cannot be ruled out, but normally there is little growth or grazing during the winter months. No toxic effects of *P. polylepis* were observed. The absence of toxicity may be because the alternate stage appears to be non-toxic or only slightly toxic (John et al. 2002).

4.1.3. Cold-water and sea-ice dinoflagellate Heterocapsa arctica subsp. frigida (III, this thesis)

Since the late 1970's, a dinoflagellate not fitting in any species description had been found in spring samples from Tvärminne area (e.g. Huttunen & Niemi 1986, Autio et al. 1990). This dinoflagellate was described as *Heterocapsa arctica* subsp. *frigida* (**III**).

Heterocapsa arctica subsp. *frigida* was mostly found in the northern parts of the investigated Baltic Sea area (the ship route between Germany and Finland; Fig. 1), overlapping with the area most probably covered in ice. The observations of the subspecies were mainly from the cold-water period (Fig. 10 in **III**). The subspecies appears to prefer areas with sea-ice cover and cold water.

Heterocapsa arctica subsp. *frigida* had the typical thecal plate arrangement for the genus *Heterocapsa*: Po, cp, 5', 3a, 6–7'', 6–7c, 5s, 5''', 2'''' (Fig. 11). It had a very distinct overall shape not resembling any other species of the genus. Examination by scanning electron microscopy revealed two types of scales (Fig. 3k in **III**) that were very similar to the scales of *H. arctica*. The extracellular body scales are considered a good character to separate different *Heterocapsa* species (Hansen 1995a). The novel dinoflagellate was smaller than *H. arctica* (Fig. 12; $\chi^2 = 278.421$, df = 3, $P < 0.001$).

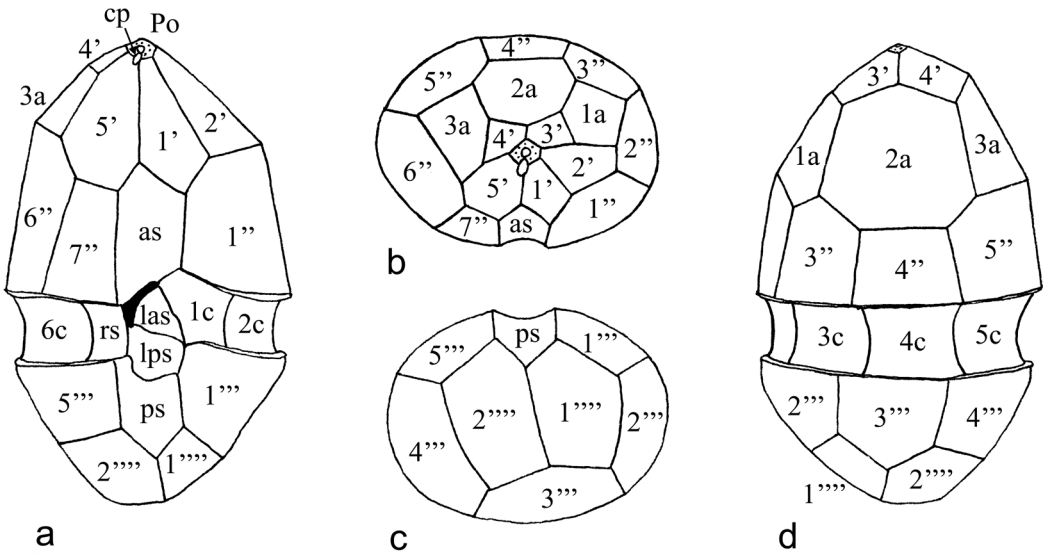


Fig. 11. The tabulation of *Heterocapsa arctica* subsp. *frigida*. (a) Ventral view, (b) apical view, (c) antapical view, and (d) dorsal view.

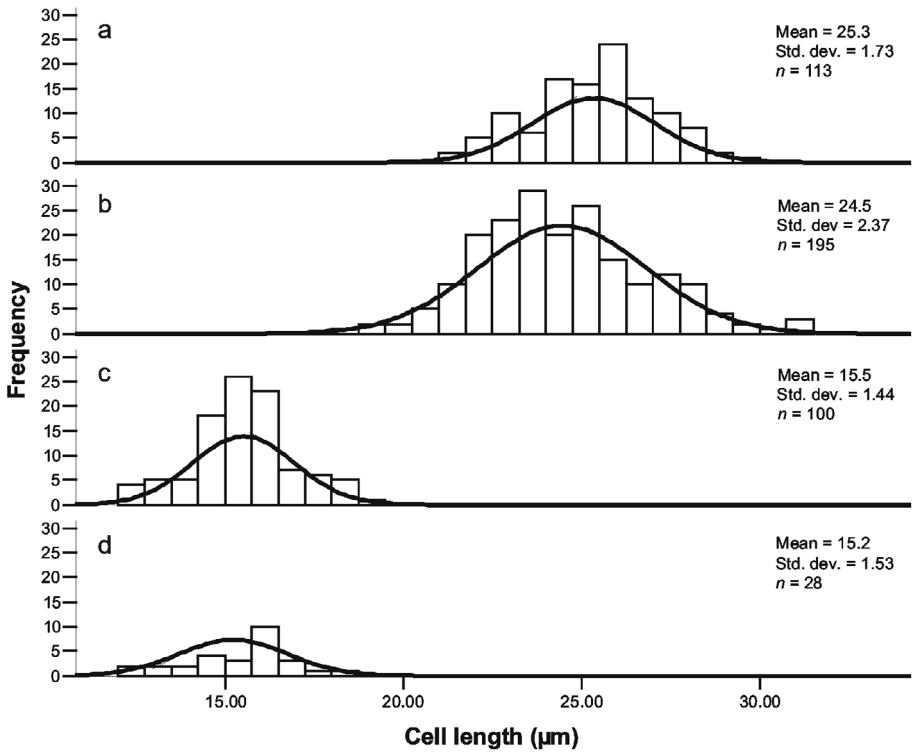


Fig. 12. Cell length of **a**) *Heterocapsa arctica* subsp. *arctica* in salinity of 35 in culture. **b**) Cell length of subsp. *arctica* in salinity of 6 in culture. **c**) Cell length of *H. arctica* subsp. *frigida* in salinity of 6 in culture. **d**) Cell length of subsp. *frigida* in natural samples.

Molecular phylogenetic comparisons were made to clarify the identity of the isolate. The 18S rRNA gene of the novel dinoflagellate and *H. arctica* were identical revealing a close relatedness of the taxa. At the time of the publication of paper **III**, only a few intraspecific copies of the ITS region sequences of *Heterocapsa* species were available in the GenBank. We interpreted the variation in the ITS region of the novel dinoflagellate and *H. arctica* was intraspecific (one base pair difference between the novel dinoflagellate and *H. arctica*, and one base pair difference between the two novel dinoflagellate clones).

Together the morphological and molecular evidence gave good grounds for describing the distinctive form of dinoflagellate as a subspecies of *H. arctica*. Some other investigators have regarded dinoflagellates that have identical ITS region sequences as different species (Logares et al. 2007), and not as same species or subspecies (i.e. regional variants, Du Rietz 1930). Logares et al. (2007) based their differentiation of *Scrippsiella hangoei* and *Peridinium aciculiferum* on differences in phenotype and salinity tolerance together with genetic isolation that was indicated by amplified fragment length polymorphism. They do not speculate the possibility of considering subspecies instead of different species.

Luckily, more sequence data have become available in GenBank after the publication of the paper **III**. This allows me to reanalyze the ITS region data in **this thesis**. One or two bases were different in the ITS1 region and three bases different (with one ambiguous position) and 5 indels in the more variable ITS2 region between the two subspecies. The minimum single base difference in the ITS1 region gives 0.19–2.07 % sequence difference between the subspecies. This difference is larger than the intraspecific divergence in other available *Heterocapsa* species sequences in GenBank (0–1.15 %), but lower than the interspecific difference (6–30 %; Fig. 13). Hence, the status of *Heterocapsa arctica* subsp. *frigida* as a subspecies of *Heterocapsa arctica* remains appropriate.

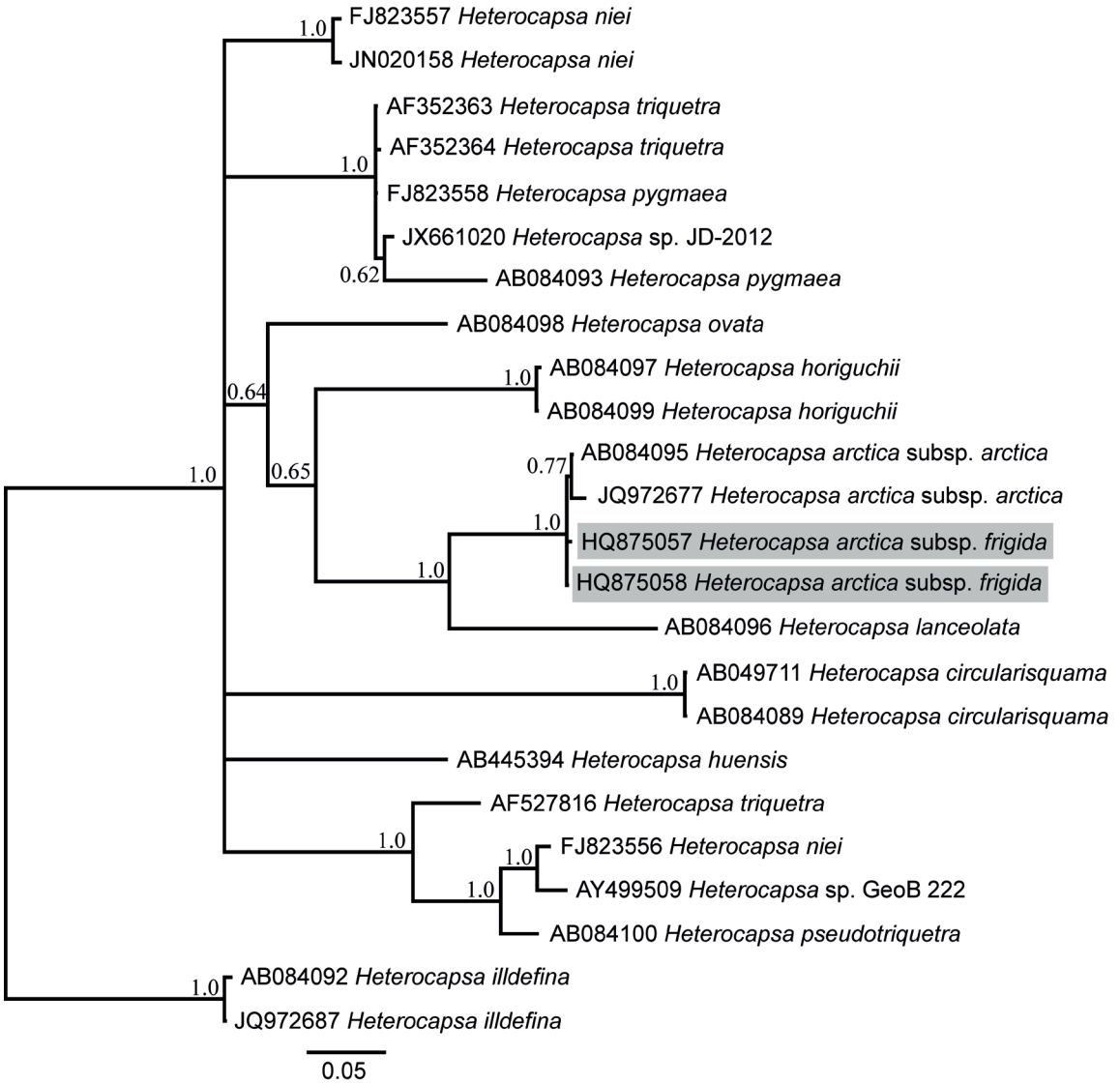


Fig. 13. Consensus Bayesian tree based on ITS region data of the members of the genus *Heterocapsa*. The sequences are denoted with the GenBank accession numbers. The sequences of *Heterocapsa arctica* subsp. *frigida* clones are highlighted with gray background. Posterior probabilities are shown near the internal nodes. Posterior probabilities were calculated across the GTR model space.

4.2. Protistan community composition of Baltic Sea ice and estimates of richness and diversity of different ice types (IV-V)

The second part of my thesis includes two environmental 18S-rRNA-gene sequencing studies. The aim of these studies was to describe the richness of higher rank protistan groups in several Baltic Sea ice and wintertime water samples (**IV-V**), and compare the richness and diversity estimates of different developmental stages (**IV**) and types (**V**) of ice. In the paper **IV**, analyses were based on 18S-rRNA-gene clone libraries, while in the paper **V**, analyses were based on 454 (Roche) pyrosequenced data. Here, I present the results of these two papers.

4.2.1. Baltic Sea ice and water community composition, using 18S-rRNA-gene clone libraries (IV)

The eight sea-ice and water samples (taken from the Gulf of Bothnia, Hanko area and outside Helsinki, samples 1–8 in Fig. 1) included 152 OTUs at a 97 % similarity level.

The supergroup Sar that includes Stramenopiles, Alveolata and Rhizaria, was the richest component of the protistan community with 108 OTUs. There were 32 ciliate, 22 dinoflagellate, 17 cercozoan, 16 diatom and 21 other stramenopile OTUs (Fig. 14). Other supergroups present in the samples were Archaeplastida (9 chlorophyte OTUs) and Opisthokonta (9 Fungi, 5 Choanomonada and 2 Metazoa OTUs). In addition, there were 6 cryptomonad, 5 unknown eukaryote, 4 kathablepharid, 2 *Telonema* and 2 haptophyte OTUs.

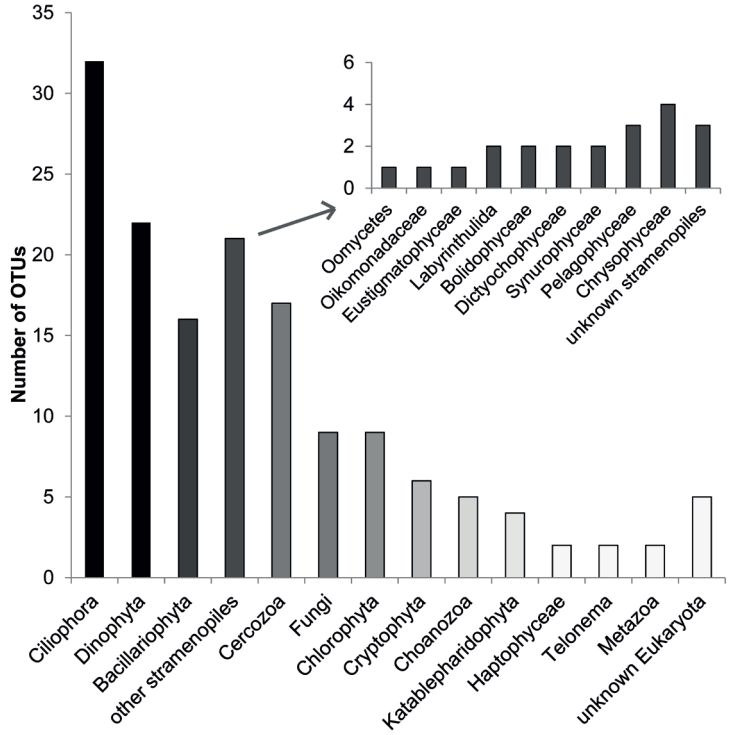


Fig. 14. Number of 97 % OTUs in the different taxonomic groups, determined using 18S rRNA gene clone libraries.

The community composition based on the clone libraries was different than that based on light microscopic examination: diatoms were not as rich as expected and ciliate, flagellate and fungi richness was surprisingly high. Firstly, lack of taxonomic rigor in identification of the latter groups in earlier sea-ice studies is the most obvious explanation for this discrepancy. Secondly, the variability of the rRNA gene copy number in different lineages (Prokopowich et al. 2003, Zhu et al. 2005) and primer annealing bias (Amacher et al. 2011) likely influenced the clone-library results. The clone libraries revealed parasitic members of the dinoflagellate group Syndiniales that are usually missed in light microscopy. Also, the DNA-based approach was very effective in differentiating similar-looking but phylogenetically different members of Fungi, Cercozoa, many Stramenopiles and Ciliophora.

The individual ice samples taken from a well-developed ice field (winter ice) as well as the pancake-ice sample had consistently higher OTU numbers and high Shannon and Simpson's diversity indices compared to the water samples (Table 3 in **IV**). The OTU-incidence and OTU-abundance based measures suggest that ice contains higher diversity of eukaryotes than water. For example, different choanomonad and ciliate OTUs were found in water (3 choanomonad, 7 ciliate OTUs) than in ice (2 choanomonad, 19 ciliate OTUs). The unweighted UniFrac measure (score = 0.611, $P < 0.001$) showed that water and the different developmental stages of sea ice (pancake, winter and spring ice) had different community composition (Fig. 15a). Although comparisons were made with equal sample sizes, the higher OTU richness in ice probably stems from the greater sampling effort in the more divergent ice.

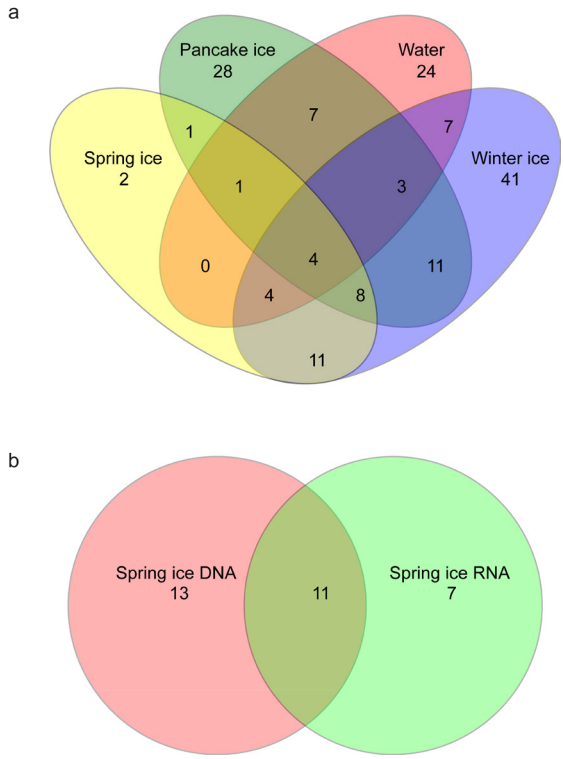


Fig 15. Number of shared 97 % OTUs in sea-ice and water samples. **a**) Venn diagram of the water, pancake ice, winter ice and spring ice sample types. The size of each oval is not proportional to the number of OTUs in each sample type. **b**) Venn diagram of DNA sample 7 and RNA sample 8. The size of the circles is not proportional to the number of OTUs in the sample.

The view was complicated when the sequence-divergence based measures were taken into account. These showed that the water sample 2 had the highest diversity of all samples (Table 3 in **IV**). So while ice had higher OTU richness compared to water, water had more phylogenetically divergent lineages. From this we may conclude that ice may be a very heterogenic environment with many niches to support high OTU richness, but not all lineages that are present in water can thrive in the ice. Twenty-four OTUs were found only in water.

During sea ice formation, particles are trapped within the ice (Reimnitz et al. 1993). Since DNA degrades slowly (e.g. Gilbert et al. 2005) the extracted DNA from sea ice may come from both living and dead organisms. In contrast, RNA is quickly degraded by intracellular enzymes (Houseley & Tollervey 2009), and RNA of dead organisms disappears from the samples fast.

In order to detect the living and active part of the ice community with molecular methods, we compared 18S rRNA and rDNA clone libraries constructed from one ice sample. Although the RNA and DNA samples had different community structure (unweighted UniFrac score = 0.671, $P < 0.001$; Fig. 15b) and the number of OTUs and diversity of the RNA sample was the lowest of all samples (Table 3 in **IV**), 16 autotrophic and heterotrophic OTUs were present in the RNA sample indicating presence of active members of the food web in the ice.

4.2.2. Baltic Sea ice and water community composition, using 454 sequencing of the 18S rRNA gene (V)

The 16 samples collected from the Gulf of Finland in March 2010 (sampling stations drift ice, pack ice and fast ice in Fig. 1) revealed 1 947 OTUs at a 97 % similarity level.

Sar was the OTU-richest supergroup including 1 576 OTUs (Fig. 16). Of these, 509 belonged to the alveolates (266 Ciliophora, 204 Dinoflagellata), 443 to stramenopiles (226 Diatomea), and 562 to Rhizaria (of which 523 were Cercozoa). The supergroup Archaeplastida included 83 OTUs (with 80 chlorophytes). Of the less certainly placed eukaryotes, the Cryptomonadales included 70 OTUs, Kathablepharidae 7 OTUs, Haptophyta 27 OTUs and *Telonema* 14 OTUs. The supergroup Opisthokonta included 91 OTUs (43 Metazoa, 33 Fungi and 15 Choanomonada). One OTU belonged to the supergroup Amoebozoa. One OTU belonged to the flagellate group Apusomonadida. Three OTUs belonged to the euglenozoan part of the supergroup Excavata. Seventy-four OTUs could not be assigned to any of the three highest rank eukaryotic lineages and therefore, they were grouped together as unknown Eukaryota.

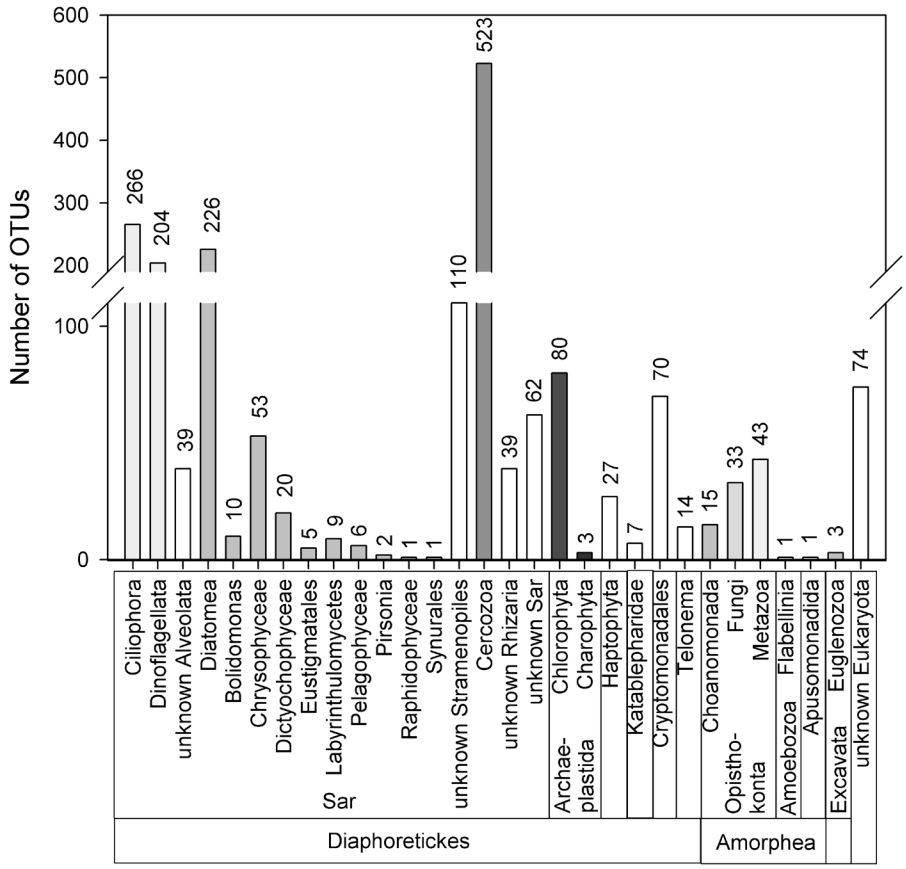


Fig. 16. Total number of OTUs in different lineages of eukaryotes in 16 Gulf of Finland ice samples. White bars denote groups with uncertain placement in the classification of Adl et al. (2012).

The overall community composition was similar to the earlier clone-library-based study (**IV**) where ciliates, dinoflagellates, diatoms and cercozoans were the richest groups in descending order. However, in paper **V**, cercozoa were the richest group followed by ciliates, diatoms and dinoflagellates. Although the possible primer bias (Amacher et al. 2011) may skew our results towards different protistan groups than in the paper **IV**, the bias does not change the fact that we found 523 cercozoan OTUs while our light microscopic analysis revealed only an all-inclusive group of 'other flagellates'.

The results of the papers **IV** and **V** highlight the richness of Cercozoa and Ciliophora in the Baltic Sea ice. These groups have mostly been ignored in earlier light-microscopy-based studies (reviewed in Granskog et al. 2006) where organisms have been counted under light microscopes using fixed samples. Heterotrophic taxa have been usually identified just to 'heterotrophic flagellates' and 'ciliates', without any information on how many different taxa are included in each group. The lack of information is probably because species-level identification of members of these

groups by light microscopy is tedious and in many cases impossible without proper staining method or expertise (e.g. Vørs 1992, Petz et al. 1995, Ikävalko 1998).

There is no surface or bottom ice community that can be found across the Gulf of Finland but there is a difference of communities in different ice types (drift, pack and fast ice). This was evident because the community membership (AMOVA Fs = 1.29, $p < 0.001$, membership-based Chi-squared distance) and composition (AMOVA Fs = 1.93, $p < 0.001$, composition-based Chi-squared distance) were different in the different ice types. This agrees with the findings of Piiparinen (2011) who showed that although the overall protistan biomass was equal in Baltic fast and drift ice samples, the ice types grouped separately in discriminant analysis because of differences in chlorophyte and heterotrophic protistan biomasses. In our study, fast ice included the most OTUs, and the drift, pack and fast ice samples shared 19 % of the OTUs (Fig. 17). This discrepancy of the assemblages of different ice types depends on the developmental history of ice, nutrient status and fine-scale ice structure (e.g. Rintala et al. 2010).

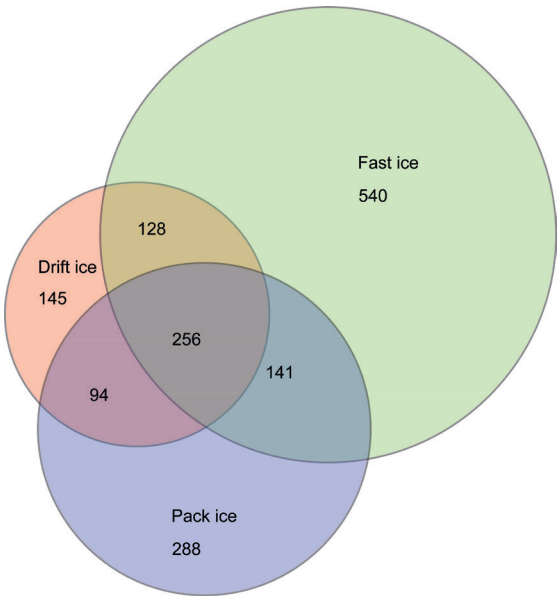


Fig. 17. Number of shared OTUs in different Gulf of Finland ice types. The size of the circles is proportional to the number of OTUs in each ice type.

4.3. Concluding remarks – linking taxonomy and environmental sequencing

I have shown that the newly collected cryptomonad is a new species, *Rhinomonas nottbecki*. The species description was based on morphological characters distinguished by light- and electron- microscopy together with molecular evidence from parts of the ribosomal gene operon (**I**). The same approach was applied to the identification of the alternate stage *Prymnesium polylepis*, which bloomed in the whole Baltic Proper during autumn–spring 2007–2008 (**II**). Also in **this thesis**, I confirmed the conclusion of the paper **III** that the cultured dinoflagellate is a novel sub-species, *Heterocapsa arctica* subsp. *frigida*.

I used environmental 18S-rRNA-gene sequencing to show that the richest eukaryotic lineages inhabiting the Baltic Sea ice are ciliates, cercozoa, dinoflagellates and diatoms (**IV-V**), all members of the supergroup Sar. Moreover, I showed that the different developmental stages (**IV**) and types (**V**) of ice have different community composition and that OTU-richness may be higher in ice than water even though water includes more divergent lineages (**IV**). I also showed that the Baltic fast ice has higher richness than pack and drift ice (**V**).

In this concluding part of my thesis, I use the taxonomic information gathered

in the first three papers to evaluate the environmental-sequencing approach of the two latter papers.

The family Pyrenomonadaceae includes three genera and 21 taxonomically accepted species (Guiry & Guiry 2013) and our ribosomal data set included 15 distinct clades (Figs. 5 and 6). The genus *Prymnesium* includes 22 accepted species (Guiry & Guiry 2013) and our 18S-rRNA-gene data set included 15 distinct branches (Fig. 10). The interspecies variability of the nuclear 18S rRNA gene was 0–2.8 % and 1–4 % in the family Pyrenomonadaceae and the genus *Prymnesium*, respectively.

When the data sets were analyzed using the program mothur and OTUs were defined at a 97 % similarity level, only one Pyrenomonadaceae and two *Prymnesium* OTUs were found. If the OTUs were defined at a 98 % similarity level, four Pyrenomonadaceae and *Prymnesium* OTUs were found. At a 99 % similarity level, eight Pyrenomonadaceae and nine *Prymnesium* OTUs were found. Use of a higher than 97 % similarity levels in the environmental sequencing studies seems justifiable based on these small tests.

Reanalysis of the data in papers **IV** and **V** revealed that if OTUs were defined at a 98 % similarity level, number of OTUs increased by 18 % to 179 OTUs in the paper **IV** and by 27 % to 2 465 OTUs in the paper **V**. At a 99 % similarity level the increase was 74 % to 264 OTUs (**IV**) and 70 % to 3 318 OTUs (**V**). Consequently, the environmental-sequencing approach used in the papers **IV** and **V** will be giving conservative estimates of the protistan richness in the Baltic Sea ice. However, errors in sequencing and alignment (Kunin et al. 2010, Bachy et al. 2013) make the use of higher similarity levels in the OTU definition questionable, and as discussed above, the variability in the 18S rDNA is not equal within different eukaryotic lineages (Caron et al. 2009). Therefore, I have used the 97 % similarity level.

The morphological examination revealed important ecological information about the blooming *Prymnesium polylepis*: the bloom was composed of the non-toxic alternate stage. Similarly, the two subspecies of *Heterocapsa arctica* are easily distinguished based on their size although they have identical 18S-rRNA-gene sequences and evidently belong to the same OTU if found in environmental sequencing studies. We can conclude that lack of taxonomic detail is not restricted to the light-microscopic surveys but is also a result of the environmental-sequencing approach.

Here, I have shown that there remains novelty to be described in the Baltic Sea, and what we know about the protistan community in Baltic Sea ice now is very incomplete. My overall conclusion is that we need to apply all available techniques, as each one provides a biased perspective on nature. Although the environmental sequencing produces data that meet the requirements of calculation of comparable diversity indices (all taxa defined at the same level), reveals cryptic taxa, and gives higher protistan richness than basic light microscopy of fixed samples, a labour intensive taxonomic approach that includes the study of live cells by light microscopy, detailed morphological description based on electron microscopy and phylogenetic analysis of suitable genetic markers gives us the best chance of finding out how many different species of protists live within the Baltic Sea ice or any other environment, and what they do.

5. ACKNOWLEDGEMENTS

I am very honored and grateful to Tove Gabrielsen for her commitment to act as my opponent. I had the pleasure to have Hannu Lehtonen to act as my kustos and I am grateful to him taking care of the official arrangements. The pre-examiners, David Caron and Rebecca Gast, gave constructive and very positive feed-back and are acknowledged for their work. The Walter and Andrée de Nottbeck Foundation has supported my work and Carl-Adam Hæggström expertly edited my thesis. Thank you.

Many thanks to my foremost supervisor, Jaanika Blomster, for guidance, pushes to the correct course and the very positive working atmosphere. My second supervisor, Harri Kuosa, gave the invaluable another perspective to my manuscripts, thank you for that. In addition to the two official supervisors, I consider Riitta Autio being my supervisor. Without her, I would not have done this thesis.

From the beginning, Janne-Markus Rintala has been the closest fellow worker. Thank you for sharing offices, ideas, travels, advice and time with me. During the final year and a half, I had the pleasure to get to know David Patterson. Thank you for the hours by the microscope and advice on professional and outside-academia life. Also, many thanks for your criticism in my thesis and manuscripts, and for checking the language.

The remainder members of the Sea Ice Ecology project: Eeva Eronen-Rasimus, Jonna Piiparinen, Susann Müller, Hermanni Kaartokallio, Anssi Vähätalo, Elina Leskinen and Jorma Kuparinen are acknowledged for the co-operation and great deal of fun during these years.

I have had the pleasure to have plenty of co-authors. I would like to thank them all equally but with special thanks to Outi Setälä for the ciliate enthusiasm, travels and time in Tvärminne and introducing me to some notable scientists, Anke Kremp for discussions and guidance, Heidi Hällfors for the excitement and companionship, Seija Hällfors for ideas and guidance and Susanna Hajdu for constructive criticism.

The colleagues and staff in the Finnish Environment Institute, at the Tvärminne Zoological Station and the Molecular Ecology and Systematics Laboratory of the University of Helsinki made my work possible, thank you all. Special thanks to Maria Piisilä and Pia Tahvanainen for teaching me the molecular work. Also, the knowledge and friends I gained from the international courses in Barcelona, Vienna, Turku, Woods Hole, Cesky Krumlov have been significant for this thesis.

I would like to thank (with some reservations) the extended Vaasa and Kallio Gangs for the dangerous, but many times needed, trips to oblivion.

My family has supported me unconditionally, thank you. Greatest thanks go to my wife, Sanna (and the boys), for sharing life with me. Hail, life!

6. REFERENCES

Adl, S. M., Simpson, A. G. B., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., James, T. Y., Karpov, S., Kugrens, P., Krug, J., Lane, C. E., Lewis, L. A., Lodge, J., Lynn, D. H., Mann, D. G., Mccourt, R. M., Mendoza, L., Moestrup, O., Mozley-Standridge, S. E., Nerad, T. A., Shearer, C. A., Smirnov, A.V., Spiegel, F. W. & Taylor, M. F. J. R. 2005: The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. – Journal of Eukaryotic Microbiology 52: 399-451.

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukes, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. A. D., Mozley-Stanridge, S. E., Pafrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A. & Spiegel, F. W. 2012: The revised classification of eukaryotes. – Journal of Eukaryotic Microbiology 59: 429-493.

Allaby, M. 2010: A Dictionary of Ecology. Fourth Edition. – Oxford University Press, Oxford, 480 pp.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. 1997: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. – Nucleic Acids Research 25: 3389-3402.

Amacher, J. A., Baysinger, C. W. & Neuer, S. 2011: The importance of organism density and co-occurring organisms in biases associated with molecular studies of marine protist diversity. – Journal of Plankton Research 33: 1762-1766.

Andersson, L. 1990: The driving force: Species concepts and ecology. – Taxon 39: 375-382.

Autio, R., Heiskanen, A.-S., Hällfors, G., Hällfors, S., Kaitala, S., Kivi, K., Kuosa, H., Kuparinen, J., Kuuppo-Leinikki, P., Lignell, R. R., Lindquist, K., Pajuniemi, R., Ranta, E., Tamminen, T. & Uitto, A. 1990: Ecological Plankton Research of the Baltic Sea. Project PELAG II Final Report 1987–1989. – PELAG Press, Helsinki. 171 pp. + 4 App.

Bachy, C., Dolan, J. R., López-García, P., Deschamps, P. & Moreira D. 2013: Accuracy of protist diversity assessments: morphology compared with cloning and direct pyrosequencing of 18S rRNA genes and ITS regions using the conspicuous tintinnid ciliates as a case study. – The ISME Journal 7: 244-255.

Björck, S. 1995: A review of the history of the Baltic Sea, 13.0–8.0 ka BP. – Quaternary International 27: 19-40.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. & Abebe, E. 2005: Defining operational taxonomic units using DNA barcode data. – Philosophical Transactions of the Royal Society of London B Biological Sciences 360: 1935-1943.

Boenigk, J., Ereshefsky, M., Hoef-Emden, K., Mallet, J. & Bass D. 2012: Concepts in protistology: Species definitions and boundaries. – European Journal of Protistology 48: 96-102.

Burki, F., Inagaki, Y., Bråte, J., Archibald, J. M., Keeling, P. J., Cavalier-Smith, T., Sakaguchi, M., Hashimoto, T., Horak, A., Kumar, S., Klaveness, D., Jakobsen, K. S., Pawlowski, J. & Shalchian-Tabrizi, K. 2009: Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, Telonemia and Centroheliozoa, are related to photosynthetic chromalveolates. – Genome Biology and Evolution 9: 231-238.

Burki, F., Okamoto, N., Pombert, J. F. & Keeling, P. J. 2012: The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. – Proceedings of the Royal Society B 279: 2246-2254.

Bütschli, O. 1880–1889: Protozoa. Palaeontologische Entwicklung der Rhizopoda. – C. F. Winter, Leipzig, 2035 pp.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters,

W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. & Knight, R. 2010: QIIME allows analysis of high-throughput community sequencing data. – Nature Methods 7: 335-336.

Caron, D. A., Countway, P. D., Savai, P., Gast, R. J., Schnetzer, A., Moorthi, S. D., Dennett, M. R., Moran, D. M. & Jones, A. C. 2009: Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. – Applied and Environmental Microbiology 75:5797-5808.

Chao, A. 1987: Estimating the population size for capture-recapture data with unequal catchability. – Biometrics 43: 783-791.

Chen, K. & Pachter, L. 2005: Bioinformatics for whole-genome shotgun sequencing of microbial communities. – PLoS Computational Biology 1: e24.

Clay, B. L., Kugrens, P. & Lee, R. E. 1999: A revised classification of Cryptophyta. – Botanical Journal of the Linnean Society 131: 131-151.

Cole, J. R., Chai, B., Marsh, T. L., Farris, R. J., Wang, Q., Kulam, S. A., Chandra, S., McGarrell, D. M., Schmidt, T. M., Garrity, G. M. & Tiedje, J. M. 2003: The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. – Nucleic Acids Research 31: 442-443.

Colwell, R. K. 2005: EstimateS: statistical estimation of species richness and shared species from samples. Version 7.5. – http://purl.oclc.org/estimates. Accessed 21 April 2011.

Darwin, C. 1859: On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. – John Murray, London, 502 pp.

Deerinck, T. J., Bushong, E. A., Thor, A. & Ellisman, M. H. 2010: NCMIR methods for 3D EM: a new protocol for preparation of biological specimens for serial block face scanning electron microscopy. – Available at: http://ncmir.ucsd.edu/sbfsem-protocol.pdf.

Denk, W. & Horstmann, H. 2004: Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. – PLoS Biology 2: e329.

Dujardin, F. 1841: Histoire naturelle des zoophytes. Infusoires, comprenant la physiologie et la classification de ces animaux, et la manière de les étudier à l'aide du microscope. – Roret, Paris, 684 pp.

Du Rietz, G. E. 1930: The fundamental units of biological taxonomy. – Svensk Botanisk Tidskrift 24: 333-428.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. 2011: UCHIME improves sensitivity and speed of chimera detection. – Bioinformatics 27: 2194-2200.

Edvardsen, B., Eikrem, W., Throndsen, J., Sáez, A. G., Probert, I. & Medlin, L. K. 2011: Ribosomal DNA phylogenies and a morphological revision provide the basis for a revised taxonomy of the Prymnesiales (Haptophyta). – European Journal of Phycology 46: 202-228.

Edgcomb, V. P., Roger, A. J., Simpson, A. G. B., Kysela, D. T. & Sogin, M. L. 2001: Evolutionary relationships among "jakobid" flagellates as indicated by alpha- and beta-tubulin phylogenies. – Molecular Biology and Evolution 18: 514-522.

Ehrenberg, C.G. 1830: Organisation, Systematik und geographisches Verhältniss der Infusionsthierchen. Zwei Vorträge, in der Akademie der Wissenschaften zu Berlin gehalten in den Jahren 1828 und 1830. – Druckerei der Königlichen Akademie der Wissenschaften, Berlin, 108 pp.

Eicken, H., Lensu, M., Leppäranta, M., Tucker, W. B., Gow, A. J. & Salmela, O. 1995: Thickenss, structure, and properties of level summer multi year ice in the Eurasian sector of the Arctic Ocean. – Journal of Geophysical Research 100: 22697-22710.

Ereshefsky, M. 2011: Mystery of mysteries: Darwin and the species problem. – Cladistics 27: 67-79.

Excoffier, L., Laval, G. & Schneider, S. 2005: Arlequin (version 3.0): an integrated software for population genetics data analysis. – Evolutionary Bioinformatics Online 1: 47-50.

Faith, D. P. 1992: Conservation evaluation and phylogenetic diversity. – Biological Conservation 61: 1-10.

Felsenstein, J. 1985: Confidence limits on phylogenies: An approach using the bootstrap. – Evolution 39: 783-791.

Felsenstein, J. 2004: PHYLIP (Phylogeny Inference Package) version 3.6. – Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Fritz, L. & Triemer, R. E. 1985: A rapid simple technique utilizing calcofluor white M2R for the visualization of dinoflagellate thecal plates. – Journal of Phycology 21: 662-664.

Garrison, R. & Buck, K. R. 1986: Organism losses during ice melting: a serious bias in sea-ice community studies. – Polar Biology 6: 237-239.

Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M. & Barnes, I. 2005: Assessing ancient DNA studies. – Trends in Ecology and Evolution 20: 541-544.

Goloboff, P., Farris, S. & Nixon, K. 2003: TNT: Tree Analysis using New Technology. – Program and documentation, available from the authors, and at www.zmuc.dk/public/phylogeny.

Granéli, E., Edvardsen, B., Roelke, D. L. & Hagström, J. A. 2012: The ecophysiology and bloom dynamics of *Prymnesium* spp. – Harmful Algae 14: 260-270.

Grafström, T. 2008: Till slut viss istillväxt i Bottenviken. – Väder och vatten 2/2008: 8.

Granskog, M., Kaartokallio, H., Kuosa, H., Thomas, D. N. & Vainio, J. 2006: Sea ice in the Baltic Sea – a review. – Estuarine, Coastal and Shelf Science 70:145-160.

Guillard, R. R. L. 1975: Culture of phytoplankton for feeding marine invertebrates. – In: Smith, W. L. & Chanley, M. H. (eds.), Culture of marine invertebrate animals, pp. 26-60. Plenum Press, New York.

Guiry, M. D. & Guiry, G. M. 2013: *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. – http://www.algaebase.org; searched on 08 February 2013.

Haeckel, E. 1866: Generelle Morphologie der Organismen. Allgemeine Grundzüge der organischen Formen-Wissenschaft, mechanisch begründet durch die von Charles Darwin reformirte Descendenz-Theorie. – G. Reimer, Berlin, 574 pp.

Haecky, P. & Andersson, A. 1999: Primary and bacterial production in sea ice in the northern Baltic Sea. – Aquatic Microbial Ecology 20: 107-118.

Haecky, P., Jonsson, S. & Andersson, A. 1998: Influence of sea ice on the composition of the spring phytoplankton bloom in the northern Baltic Sea. – Polar Biology 20: 1-8.

Hällfors, G. & Niemi, Å. 1974: A *Chrysochromulina* (Haptophyceae) bloom under the ice in the Tvärminne archipelago, southern coast of Finland. – Memoranda Societatis pro Fauna et Flora Fennica 50: 89-104.

Hällfors, G. & Thomsen, H. A. 1979: Further observations on *Chrysochromulina birgeri* (Prymnesiophyceae) from the Tvärminne archipelago, SW coast of Finland. – Acta Botanica Fennica 110: 41-46.

Hansen, G. 1995a: Analysis of the thecal plate pattern in the dinoflagellate *Heterocapsa rotundata* (Lohmann) comb. nov. (=*Katodinium rotundatum* (Lohmann) Loeblich). – Phycologia 34: 166-170.

Hansen, G. 1995b: The use of scanning electron microscopy in dinoflagellate taxonomy. – Scanning 17: suppl. 97.

Hansson, M. & Andersson, L. 2008: Cruise report from R/V Argos. – Swedish Meteorological and Hydrological Institute Oceanographical Laboratory, Dnr: Mo-2008-025. Available at: http://www.smhi.se/polopoly_fs/1.6438!exp_0808eng.pdf.

Harper, J. L. & Hawksworth, D. L. 1995: Preface. – In: Hawksworth, D. L. (ed.), Biodiversity; measurement and estimation, pp. 5-12. Chapman & Hall, London.

Hausdorf, B. 2011: Progress toward a general species concept. – Evolution 65: 923-931.

Hausmann, K., Hülsmann, N. & Radek, R. 2003: Protistology. 3rd completely revised edition. – E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 379 pp.

HELCOM 2008: Manual for marine monitoring in the COMBINE programme of HELCOM. – Available at http://www.helcom.fi/groups/monas/CombineManual/en_GB/Contents/.

Hellström, S. 2008: Rekordvarm vinter i östra Sverige. – Väder och vatten 2/2008: 11.

Hill, D. R. A. 1991: A revised circumscription of *Cryptomonas* (Cryptophyceae) based on examination of Australian strains. – Phycologia 30: 170-188.

Hill, D. R. A. & Wetherbee, R. 1988: The structure and taxonomy of *Rhinomonas pauca* gen et sp. nov. (Cryptophyceae). – Phycologia 27: 355-365.

Hill, D. R. A. & Wetherbee, R. 1989: A reappraisal of the genus *Rhodomonas* (Cryptophyceae). – Phycologia 28: 143-158.

Hoef-Emden, K. 2007: Revision of the genus *Cryptomonas* (Cryptophyceae) II: incongruences between the classical morphospecies concept and molecular phylogeny in smaller pyrenoid-less cells. – Phycologia 46: 402-428.

Hoef-Emden, K., Marin, B. & Melkonian, M. 2002: Nuclear and nucleomorph SSU rDNA phylogeny in the Cryptophyta and the evolution of cryptophyte diversity. – Journal of Molecular Evolution 55: 161-179.

Hoef-Emden, K. & Melkonian, M. 2003: Revision of the genus *Cryptomonas* (Cryptophyceae): a combination of molecular phylogeny and morphology provides insights into a long-hidden dimorphism. – Protist 154: 371-409.

Houseley, J. & Tollervey, D. 2009: The many pathways of RNA degradation. – Cell 136: 763-776.

Huber, T., Faulkner, G. & Hugenholtz, P. 2004: Bellerophon; a program to detect chimeric sequences in multiple sequence alignments. – Bioinformatics 20: 2317-2319.

Huelsenbeck, J. P., Larget, B., Miller, R. E. & Ronquist, F. 2002: Potential applications and pitfalls of Bayesian inference of phylogeny. – Systematic Biology 51: 673-688.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. 2001: Bayesian inference of phylogeny and its impact on evolutionary biology. – Science 294: 2310-2314.

Huttunen, M. & Niemi, Å. 1986: Sea-ice algae in the Northern Baltic Sea. – Memoranda Societatis pro Fauna et Flora Fennica 62: 58-62.

Ikävalko, J. 1998: Further observations on flagellates within sea ice in northern Bothnian Bay, the Baltic Sea. – Polar Biology 19: 323-329.

Ikävalko, J. & Thomsen, H. A. 1996: Scale-covered and loricate flagellates (Chrysophyceae and Synurophyceae) from the Baltic Sea ice. – Nova Hedwigia, Beiheft 114: 147-160.

Ikävalko, J. & Thomsen, H. A. 1997: The Baltic Sea ice biota (March 1994): a study of the protistan community. – European Journal of Protistology 33: 229-243.

International Commission of Zoological Nomenclature 1999: International Code of Zoological Nomenclature, Glossary. – The International Trust for Zoological Nomenclature. Available at http://www.nhm.ac.uk/hosted-sites/iczn/code/?booksection=glossary&nfv=true&mF=.

Iwataki, M., Botes, L., Sawaguchi, T., Sekiguchi, K. & Fukuyo, Y. 2003: Cellular and body scale structure of *Heterocapsa ovata* sp. nov. and *Heterocapsa orientalis* sp. nov. (Peridiniales, Dinophyceae). – Phycologia 42: 629-637.

Jensen, M. Ø. 1998: The genus *Chrysochromulina* (Prymnesiophyceae) in Scandinavian coastal waters – diversity, abundance and ecology. – Ph.D. Thesis, University of Copenhagen, Denmark.

Jensen, M. Ø. & Moestrup, Ø. 1999: Ultrastructure of *Chrysochromulina ahrengotii* sp. nov. (Prymnesiophyceae), a new saddleshaped species of *Chrysochromulina* from Danish coastal waters. – Phycologia 38: 195-207.

John, U., Tillmann, U. & Medlin, L. K. 2002: A comparative approach to study inhibition of grazing and lipid composition of a toxic and non-toxic clone of *Chrysochromulina polylepis* (Prymnesiophyceae). – Harmful Algae 1: 45-57.

Judd, W.S., Campbell, C.S., Kellog, E.A., Stevens, P.F. & Donoghue, M.J. 2008: Plant Systematics: A Phylogenetic Approach, Third Edition. – Sinauer Associates, Sunderland, 611 pp.

Kaartokallio, H. 2004: Food web components, and physical and chemical properties of Baltic Sea

ice. – Marine Ecology Progress Series 273: 49-83.

Kaartokallio, H., Kuosa, H., Thomas, D. N., Granskog, M. A. & Kivi, K. 2007: Biomass, composition and activity of organism assemblages along a salinity gradient in sea ice subjected to river discharge in the Baltic Sea. – Polar Biology 30:183-197.

Katoh, K. & Toh, H. 2008: Recent developments in the MAFFT multiple sequence alignment program. – Briefings in Bioinformatics 9: 286-298.

Kottmeier, S. T. & Sullivan, C. W. 1988: Sea-ice microbial communities 9. Effects of temperature and salinity on rates of metabolism and growth of auttrophs and heterotrophs. – Polar Biology 8: 293-304.

Kunin, V., Engelbrektson, A., Ochman, H. & Hugenholtz, P. 2010: Wrinkles in the rare biosphere: pyrosequencing error can lead to artificial inflation of diversity estimates. – Environmental Microbiology 12: 118-123.

Kuosa, H., Borrman, B., Kivi, K. & Brandini, F. 1992: Effects of Antarctic sea ice biota on seeding as studied in aquarium experiments. – Polar Biology 12: 333-339.

Kuvardina, O. N., Leander, B. S., Aleshin, V. V., Mylnikov, A. P., Keeling, P. J. & Simdyanov, T. G. 2002: The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to apicomplexans. – Journal of Eukaryotic Microbiology 49: 498-504.

Lake, I. & Grafström, T. 2008: Fortsatt långsam avkylning men trots detta onormalt lite is. – Väder och vatten 1/2008: 8.

Lee, M. S. Y. 1997: Documenting present and past biodiversity: conservation biology meets palaeontology. – Trends in Ecology and Evolution 12: 132-133.

Lee, S.-H. & Chao, A. 1984: Estimating population size via sample coverage for closed capture-recapture models. – Biometrics 50: 88-97.

Leppäranta, M. & Myrberg, K. 2008: Physical oceanography of the Baltic Sea. – Springer, Berlin, 378 pp.

Linnæus, C. 1753: Species plantarum, exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas. – L. Salvius, Holmiæ, 1200 pp.

Linnæus, C. 1758: Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata. – L. Salvius, Holmiæ, 824 pp.

Logares, R., Rengefors, K., Kremp, A., Shalchian-Tabrizi, K., Boltovskoy, A., Tengs, T., Shurtleff, A. & Klaveness, D. 2007: Phenotypically different microalgal morphospecies with identical ribosomal DNA: a case of rapid adaptive evolution? – Microbial Ecology 53: 549-61.

Lowe, C. D., Keeling, P. J., Martin, L. E., Slamovits, C. H., Watts, P. C. & Montagnes, D. J. S. 2011: Who is *Oxyrrhis marina*? Morphological and phylogenetic studies on an unusual dinoflagellate. – Journal of Plankton Research 33: 555-567.

Lozupone, C. A. & Knight, R. 2008: Species divergence and the measurement of microbial diversity. – FEMS Microbiology Reviews 32: 557-578.

Maggs, C. A. & Ward, B. A. 1996: The genus *Pikea* (Dumontiaceae, Rhodophyta) in England and the North Pacific: comparative morphological, life history, and molecular studies. – Journal of Phycology 32: 176-193.

Magurran, A. E. 2004: Measuring Biological Diversity. – Blackwell, Oxford, 256 pp.

Majaneva, M., Autio, R., Huttunen, M., Kuosa, H. & Kuparinen J. 2009: Phytoplankton monitoring: the effect of sampling methods used during different stratification and bloom conditions in the Baltic Sea. – Boreal Environment Research 14: 313-322.

Martin, A. P. 2002: Phylogenetic approaches for describing and comparing the diversity of microbial communities. – Applied and Environmental Microbiology 68: 3673-3682.

Mayr, E. (ed.) 1957: The Species Problem. – American Association for the Advancement of Science, Washington D.C., 395 pp.

Meiners, K., Fehling, J., Granskog, M. A. & Spindler, M. 2002: Abundance, biomass and composition of biota in Baltic Sea ice and underlying water (March 2000). – Polar Biology 25:761-770.

Mikkelsen, D., Rysgard, S. & Glud, R. 2008: Microalgal composition and primary production in Arctic sea ice: A seasonal study from Kobbefjord (Kangerluarsunnguaq), West Greenland. – Marine Ecology Progress Series 368: 65-74.

Moestrup, Ø. & Thomsen, H. 1980: Preparation of shadow cast whole mounts. – In: Gatt, E. (ed.), Handbook of phycological methods. Developmental and cytological methods, pp. 385-390. Cambridge University press, Cambridge.

Moon-van der Staay, S.Y., De Wachter, R. & Vaulot, D. 2001: Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. – Nature 409: 607-610.

Norrman, B. & Andersson, A. 1994: Development of ice biota in a temperate sea area (Gulf of Bothnia). – Polar Biology 14: 531-537.

Not, F., Valentin, K., Romari, K., Lovejoy, C., Massana, R., Tobe, K., Vaulot, D. & Medlin, K. 2007: Picobiliphytes: a marine picoplanktonic algal group with unknown affinities to other eukaryotes. – Science 315: 253-255.

Novarino, G. 1991a: Observations on *Rhinomonas reticulata* comb. nov. and *Rhinomonas reticulata* var. *eleniana* var. nov. (Cryptophyceae), with comments on the genera *Pyrenomonas* and *Rhodomonas*. – Nordic Journal of Botany 11: 243-252.

Novarino, G. 1991b: Observations on some new and interesting Cryptophyceae. – Nordic Journal of Botany 11: 599-611.

Paasche, E., Edvardsen, B. & Eikrem, W. 1990: A possible alternate stage in the life cycle of *Chrysochromulina polylepis* Manton et Parke (Prymnesiophyceae). – Nova Hedwigia, Beiheft 100: 91-99.

Page, R. D. M. & Holmes, E. C. 1998: Molecular evolution: a phylogenetic approach. – Blackwell Science Ltd., Oxford, 346 pp.

Palosuo, E. 1961: Crystal structure of brackish and freshwater ice. – International Association of Science Hydrology 54: 9-14.

Parfrey, L. W., Grant, J., Tekle, Y. I., Lasek-Nesselquist, E., Morrison, H. C., Sogin, M. L., Patterson, D. J. & Katz, L. 2010: Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. – Systematic Biology 59: 518–533.

Patterson, D. J. 1985: On the organization and affinities of the amoeba, *Pompholyxophrys punicea* Archer, based on ultrastructural examination of individual cells from wild material. – Journal of Protozoology 32: 241-246.

Patterson, D. J. 1999: The diversity of eukaryotes. – American Naturalist 154: S96-124.

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirku, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., Lukes, J., Mann, D. G., Mitchell, E. A. D., Nitsche, F., Romeralo, M., Saunders, G. W., Simpson, A. G. B., Smirnov, A. V., Spouge, J. L., Stern, R. F., Stoeck, T., Zimmermann, J., Schindel, D. & de Vargas, C. 2012: CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. – PLoS Biology 10: e1001419.

Pereyra, R. T., Bergström, L., Kautsky, L. & Johannesson, K. 2009: Rapid speciation in a newly opened postglacial marine environment, the Baltic Sea. – BMC Evolutionary Biology 9: 70.

Petrich, C. & Eicken H. 2010: Growth, structure and properties of sea ice. – In: Thomas, D. N. & Dieckmann, G. S. (eds.), Sea ice, 2nd edition, pp. 23-77. Blackwell Publishing Ltd, Oxford.

Petz, W., Song, W. & Wilbert N 1995: Taxonomy and ecology of the ciliate fauna (Protozoa, Ciliophora) in the endopagial and pelagial of the Weddel Sea, Antarctica. – Stapfia 40: 1-223.

Piiparinen, J. 2011: Fast- and drift-ice communities in the Bothnian Bay and the impact of UVA ra-

diation on the Baltic Sea ice ecology. – Walter and Andrée de Nottbeck Foundation Scientific Reports 36: 1-57.

Piiparinen, J., Kuosa, H. & Rintala, J.-M. 2010: Winter-time ecology in the Bothnian Bay, Baltic Sea: nutrients and algae in fast ice. – Polar Biology 33: 1445-1461.

Posada, D. 2008: jModelTest: Phylogenetic Model Averaging. – Molecular Biology and Evolution 25: 1253-1256.

Prokopowich, C. D., Gregory, T. R. & Crease, T. J. 2003: The correlation between rDNA copy number and genome size in eukaryotes. – Genome 46: 48-50.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. & Glöckner, F. O. 2013: The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. – Nucleic Acids Research 41 (D1): D590-D596.

Rantajärvi, E. 2003: Alg@line in 2003: 10 years of innovative plankton monitoring and research and operational information service in the Baltic Sea. – MERI, Report Series of the Finnish Institute of Marine Research 48: 1-55.

Rantajärvi, E., Olsonen, R., Hällfors, S., Leppänen, J.-M. & Raateoja, M. 1998: Effect of sampling frequency on detection of natural variability in phytoplankton: unattended high-frequency measurements on board ferries in the Baltic Sea. – ICES Journal of Marine Science 55: 697-704.

Remane, A. 1934: Die Brackwasserfauna. – Verhandlungen der Deutschen Zoologischen Gesellschaft 36: 34-74.

Reimnitz, E., Clayton, J. R., Kempema, E. W., Payne, J. R. & Weber, W. S. 1993: Interaction of rising frazil with suspended particles: tank experiments with applications to nature. – Cold Regions Science and Technology 21: 117-135.

Rintala, J.-M., Piiparinen, J., Ehn, J., Autio, R. & Kuosa, H. 2006: Changes in phytoplankton biomass and nutrient quantities in sea ice as responses to light/dark manipulations during different phases of the Baltic winter 2003. – Hydrobiologia 554: 11-24.

Rintala, J.-M., Piiparinen, J. & Uusikivi, J. 2010: Drift-ice and under-ice water communities in the Gulf of Bothnia (Baltic Sea). – Polar Biology 33: 179-191.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. 2012: MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. – Systematic Biology 61: 539-542.

Rudenberg, G. H. & Rudenberg, P. G. 2010: Chapter 6 – Origin and Background of the Invention of the Electron Microscope: Commentary and Expanded Notes on Memoir of Reinhold Rüdenberg. – Advances in Imaging and Electron Physics 160: 207-286.

Schloss, P. D., Gevers, D. & Westcott, S. L. 2011: Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. – PLoS ONE 6: e27310.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. & Weber, C. F. 2009: Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. – Applied and Environmental Microbiology 75: 7537-7541.

Shannon, C. E. 1948: A mathematical theory of communication. – The Bell System Technical Journal 27: 379-423 and 623-656.

Simpson, E. H. 1949: Measurement of diversity. – Nature 163: 688.

Slobodchikoff, C. N. (ed.) 1976: Concepts of Species. – Dowden, Hutchinson & Ross, Stroudsburg, 368 pp.

Sogin, M. L., Elwood, H. J. & Gunderson, J. H. 1986: Evolutionary diversity of eukaryotic small-subunit rRNA genes. – Proceedings of the National Academy of Sciences of the United States of America 83: 1383-1387.

Sokal, R. R. & Sneath, P. H. A. 1963: Principles of numerical taxonomy. – W. H. Freeman & Co, San Francisco, 359 pp.

Suutari, M. Majaneva, M., Fewer, D. P., Voirin, B., Aiello, A., Friedl, T., Chiarello, A. G. & Blomster, J. 2010: Molecular evidence for a diverse green algal community growing in the hair of sloths and a specific association with *Trichophilus welckeri* (Chlorophyta, Ulvophyceae). – BMC Evolutionary Biology 10: 86.

Takishita, K., Kouichirou, N. & Uchida, A. 2000: Origin of the plastid in the anomalously pigmented dinoflagellate *Gymnodinium mikimotoi* (Gymnodiniales, Dinophyta) as inferred from phylogenetic analysis based on the gene encoding the large subunit of form I-type RuBisCO. – Phycological Research 48: 85-89.

Tamelander, T., Reigstad, M., Hop, H., Carroll, M. L. & Wassmann, P. 2008: Pelagic and sympagic contribution of organic matter to zooplankton and vertical export in the Barents Sea marginal ice zone. – Deep Sea Research part II 55: 2330-2339.

Tavaré, S. 1986: Some probabilistic and statistical problems in the analysis of DNA sequences. – In: Miura, R. M. (ed.), Some mathematical questions in biology – DNA sequence analysis: pp. 57-86. American Mathematical Society, Providence.

Thorstensson, B. & Yhlen, B. 2007: Expeditionsrapport från U/F Argos. – Swedish Meteorological and Hydrological Institute Oceanographical Laboratory, Dnr: Mo-2007-211. Available at: http://www.smhi.se/polopoly_fs/1.2426!exp_495007.pdf.

Tikkanen, M. & Oksanen, J. 2002: Late Weichselian and Holocene shore displacement history of the Baltic Sea in Finland. – Fennia 180: 9-20.

United Nations Environment Programme 1992: Convention on biological diversity, June 1992. – United Nations Environment Programme Environmental Law and Institutions Programme Activity Centre, Nairobi, 28 pp.

Utermöhl, H. 1958: Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. – Mitteilungen der Internationale Vereinigung für theoretische und angewandte Limnologie 9: 1-38.

Van de Peer, Y., Van der Auwera, G. & De Wachter, R. 1996: The evolution of stramenopiles and alveolates as derived by ”substitution rate calibration” of small ribosomal subunit RNA. – Journal of Molecular Evolution 42: 201-210.

Vørs, N. 1992: Heterotrophic amoebae, flagellates and heliozoan from the Tvärminne area, Gulf of Finland, in 1988–1990. – Ophelia 36: 1-109.

Webb, C. O., Ackerly, D. D. & Kembel, S. W. 2008: Phylocom: software for the analysis of phylogenetic community structure and trait evolution. – Bioinformatics 24: 2098-2100.

Willén, T. 1962: Studies on the phytoplankton of some lakes connected with or recently isolated from the Baltic. – Oikos 13: 169-199.

Wilson, E. O. 1988: Biodiversity. – National Academy Press, Washington, 538 pp.

Woese, C. R. & Fox, G. E. 1977: The concept of cellular evolution. – Journal of Molecular Evolution 10:1-6.

Wu, R. S. S. 1982: Effects of taxonomic uncertainty on species diversity indices. – Marine Environmental Research 6: 215-225.

Yhlen, B. & Andersson, L. 2007: Expeditionsrapport från U/F Argos. – Swedish Meteorological and Hydrological Institute Oceanographical Laboratory, Dnr: Mo-2007-192. Available at: http://www.smhi.se/polopoly_fs/1.2418!exp_404407.pdf.

Zimmermann, J., Jahn, R. & Gemeinholzer, B. 2011: Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. – Organisms Diversity & Evolution 11: 173-192.

Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. 2005: Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. – FEMS Microbiology Ecology 52: 79-92.

Zwickl, D. J. 2006: Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. – Ph.D. dissertation, The University of Texas at Austin.